GLOBAL DISTRIBUTION AND GENETICS OF SOIL YEASTS

GLOBAL DISTRIBUTION, EPIDEMIOLOGY AND GENETICS OF INFECTIOUS YEAST *CRYPTOCOCCUS DENEOFORMANS* AND OTHER SOIL YEASTS

By HIMESHI SAMARASINGHE BSc., MSc.

A Thesis Submitted to the School of Graduate Studies in the Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

Doctor of Philosophy (2021)
Department of Biology
McMaster University
Hamilton, Ontario, Canada

TITLE: Global Distribution, Epidemiology and Genetics of Infectious Yeast *Cryptococcus deneoformans* and Other Soil Yeasts

AUTHOR: Himeshi Samarasinghe, BSc., MSc. (McMaster University)
SUPERVISOR: Dr. Jianping Xu

NUMBER OF PAGES: xiv, 233

Lay Abstract

Yeast species have been an integral part of human civilization for centuries, with earliest evidence of yeast-fermented beverages dating back to 7000 B.C. Soil is a primary yeast habitat where yeasts with diverse profiles, including those infectious to humans, reside. *Cryptococcus deneoformans* is a soil yeast that infects people via the respiratory tract to cause serious systemic infections, resulting in 181,000 deaths worldwide every year. *C. deneoformans*'s ability to produce melanin pigments is a crucial factor in its success as an infectious agent. My thesis investigated global patterns in soil yeast diversity, with a focus on the genetic factors that regulate melanin synthesis in *C. deneoformans.* I explored the potential impacts of anthropogenic activities on soil yeast communities, as exemplified by the ongoing hybridization between *C. deneoformans* and its sister species, which were previously geographically isolated. Findings of my thesis advance our knowledge on yeast biology, infections, and genetics research.

Abstract

Yeasts, broadly defined as unicellular fungi, have a disproportionate impact on human health and economy despite comprising less than 1% of the fungal kingdom. Soil is a primary yeast habitat where they play essential roles in decomposition, nutrient cycling and as food sources for other soil dwellers. *Cryptococcus deneoformans* is a basidiomycetous yeast commonly found in soil in association with pigeon droppings. As an opportunistic pathogen of humans, it contributes to 181,000 deaths caused by cryptococcosis and fungal meningitis worldwide every year. Significant intraspecific variation in melanin synthesis, an essential virulence factor of *C. deneoformans*, is observed in natural populations, with its genetic basis remaining largely unknown. My thesis investigated global patterns in soil yeast diversity where we identified mean annual precipitation and international human travel as two, strong predictors of soil yeast diversity worldwide. We discovered a novel *C. deneoformans* population in Saudi Arabian desert soils, likely a recent introduction to the region facilitated by anthropogenic activities. Using bulk segregant analysis and gene expression assays, we identified six, novel candidate genes that potentially contribute to intraspecific melanin variance in *C. deneoformans.* Finally, we investigated genome-wide allele distribution patterns in hybrid strains derived from mating between *C. deneoformans* and its sister species, *C. neoformans.* Significantly skewed allele distributions we detected in hybrid genomes highlight the genomic incompatibilities between the two species and support their classification into two, distinct species which has been a topic of much debate. Overall, my PhD thesis makes several contributions to our understanding of soil yeast communities, genetics of virulence factors in *C. deneoformans,* and outcomes of hybridization between fungal lineages.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Jianping Xu, for all the academic and moral support he has given me throughout my graduate career. I truly appreciate the independence and freedom he gave me in research to follow my interests and develop my own ideas. He has always been supportive of me exploring opportunities to develop my career, both in and outside of the lab. I am thankful to Dr. Ian Dworkin for all his guidance on research, data analysis and statistics over the years and for making time for me whenever I had questions. I would like to thank my committee members Dr. Jim Quinn, and Dr. Jon Stone for critically evaluating my work, providing helpful feedback, and their continued support throughout my graduate career to see me succeed.

Xu lab members, past and present, have been an integral part of my success both in and outside of the lab. Dr. Aaron Vogan was a wonderful mentor to me when I first joined the lab and helped me get comfortable with lab work. The work I have done would not have been possible without the support of Heather Yoell, Dr. Eta Ashu, Dr. Adrian Forsythe, Sarah Sandor, Man You, Gregory Korfanty, Yuying Fan, Yue Wang, Meagan Archer, Renad Aljohani, Wenjing Hua and Kelly Dong. Yi Lu, Ahmad Al-Amad, Meghan Craughwell, Carlene Jimenez and Lisa Teng, as undergraduate thesis students or volunteers, made significant contributions to my PhD thesis work presented here.

In the fall of 2019, Dr. Joe Heitman graciously welcomed me to his lab at Duke University, North Carolina where he generously supported me in my endeavour to learn CRISPR-Cas9 gene editing, presented in chapter 4. Dr. Sheng Sun, a Xu lab alumnus and currently an assistant research professor at Duke University, was a kind and supportive mentor to me throughout my fellowship there. Additionally, I would like to thank Dr. Vikas Yadav, Dr. Ci Fu, Dr. Yufent Fang and all Heitman lab members for welcoming me to their lab, helping me with experiments and providing guidance and support.

Lastly, I would like to thank my parents, family, friends, and my partner Jenis Jesuratnam for their constant love and support. I am grateful for their encouragement and for believing in me even when I did not.

Table of Contents

List of Figures

List of Tables

Abbreviations

CNSC – *Cryptococcus neoformans* species complex

CGSC – *Cryptococcus gattii* species complex

MLST – multilocus sequence typing

ST – sequence type

YEPD – yeast extract peptone dextrose

QTL – quantitative trait loci

BSA – bulk segregant analysis

RT-qPCR – reverse transcriptase quantitative polymerase chain reaction

PCR-RFLP – polymerase chain reaction restriction fragment length polymorphism

Chapter 1

Introduction

1.1 Yeasts of the Fungal Kingdom

The kingdom of fungi comprises multicellular eukaryotic organisms such as mushrooms and molds, along with unicellular eukaryotic organisms, broadly defined as yeasts. Given the estimated 1.5 to 5 million fungal species currently in existence, yeast species, estimated to be approximately 1500 in number, make up less than 1% of the fungal kingdom (Blackwell 2011; Kurtzman, Fell, and Boekhout 2011; Wu et al. 2019). As some of the earliest domesticated organisms in human history, yeasts have maintained a close association with human health, food, and economy over the centuries. *Saccharomyces cerevisiae*, best known as brewer's and baker's yeast, is used worldwide in alcohol fermentation and baking, with the earliest evidence of fermented beverages dating back to Neolithic times in China (McGovern et al. 2004). A small number of yeast species are opportunistic human pathogens capable of causing skin infections, vaginal infections, and more serious systemic infections such as fungal meningitis. Nearly a billion people are afflicted with a fungal infection with a considerable proportion caused by pathogenic yeast species (Bongomin et al. 2017). Approximately 70% of females worldwide suffer from at least one vaginal yeast infection, caused by *Candida* species, in their lifetime (Sobel et al. 1998).

Despite the relatively small number, yeast species exhibit significant diversity in morphology, ecology, and lifecycle, and are found in both major phyla of the fungal kingdom, namely Ascomycota and Basidiomycota (Figure 1). Yeasts were first observed microscopically in 1680 by Anton van Leeuwenhoek but they were not recognized as fungi until 1837 (Barnett 2004). Historically, yeasts were classified into taxonomic groups based on morphology and physiology, but DNA sequencing has now become the norm for establishing species identity and taxonomic classification.

In 2012, the intergenic transcribed spacer (ITS) region of the small subunit nuclear ribosomal RNA gene cluster was formally recommended as the primary fungal barcode to be used in species classification (Schoch et al. 2012). It is now common practice for yeast strains isolated from clinical and environmental sources to be sequenced at the ITS locus to establish species identity. Large-scale metagenomics investigations often use the ITS region as the barcode for characterizing fungal sequences. D1/D2 domain of the large subunit (LSU) ribosomal nuclear DNA is also sometimes used as a secondary fungal barcode in species characterization, especially in yeasts (Xu 2016). While both ITS and LSU barcodes are important for characterizing species of the fungal kingdom, sequences of neither loci can be used as the only criteria for the morphological distinction between yeasts and filamentous fungi. Since yeasts do not form a monophyletic group within the fungal kingdom, barcode sequences do not have any yeast-specific sequence motifs that can be used to select for yeast sequences. Researchers rely on annotated yeast sequences deposited in public databases such as NCBI and UNITE to identify novel sequences as yeasts.

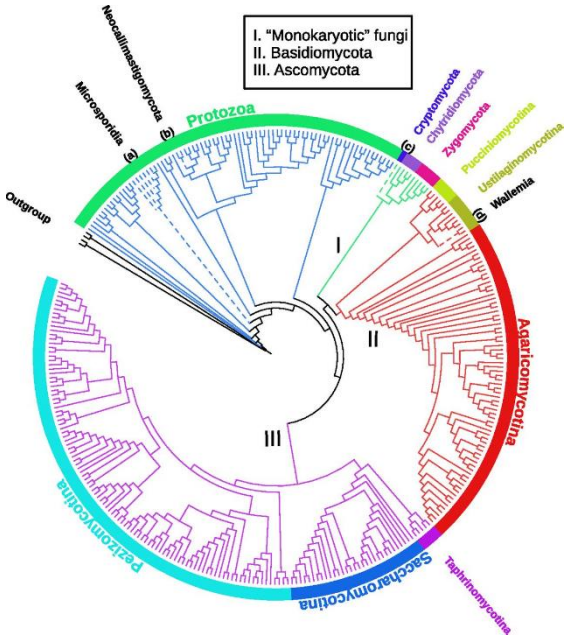


Figure 1: Phylogenetic tree of the fungal kingdom. Yeasts are found in both Basidiomycota (II) Ascomycota (III), making them taxonomically diverse as a group. Adapted from Fig 1 in Choi and Kim (2017).

1.2 Global distribution of soil yeasts

While scientists' interest in yeasts dates back to at least the 17th century, soil was not determined to be a primary yeast habitat until the 1950s (reviewed in Yurkov 2018). Due to their extensive use in fruit fermentation, yeasts were often associated with orchards, plants and rotting fruits with any yeast cells found in soil being considered transient. Pioneering studies starting in 1920s onwards that isolated and quantified yeasts from a variety of soils finally determined yeasts to be true soil-dwellers where they lived and reproduced. While yeasts rarely exceed thousands of cells per gram of soil, they form an integral part of soil ecosystems as decomposers of organic material and nutrient sources for other soil-inhabiting protists, bacteria, and nematodes (Botha 2011; Yurkov 2018). In fact, yeasts may be the predominant type of soil fungi in cold biospheres such as continental Antarctica (Connell et al. 2008; Vishniac 1996).

Soil is a primary reservoir of common pathogenic yeasts: humans can get infected via inhalation of desiccated cells or spores originating from contaminated soils (Kurtzman, Fell, and Boekhout 2011). The majority of yeast infections are caused by species of the *Candida* genus (Bongomin et al. 2017). Most pathogenic *Candida* species, with the exception of *C. albicans* which is a natural component of the human gut and oral microflora, are commonly found in soil with a global distribution (Papon et al. 2013). The non-albicans species, notably *C. tropicalis* and *C. glabrata,* are considered emerging pathogens due to their expanding distributions and antifungal resistance (Opulente et al. 2019; Yang et al. 2012). Soil is also a primary habitat of pathogenic *Cryptococcus* species, responsible for ~220 000 infections worldwide with 80% mortality. The causative agents, *C. neoformans, C. deneformans,* and *C. gattii* have been isolated from soil in all continents except Antarctica (Cogliati 2013). Recently, the environmental reservoir of the multi-drug resistant superbug *Candida auris,* responsible for hospital outbreaks in over 33 countries within the past decade, was discovered to be sediment soil and seawater in tropical coastal environments (Arora et al. 2021; Jeffery-Smith et al. 2018).

Soil yeast populations often differ in structure and composition between locations. Findings of many studies that investigated soil yeast diversity over the last few decades reveal that, apart from a few widespread genera such as *Cyberlindnera,*

Schwanniomyces, *Naganishia*, *Goffeauzyma* and *Solicoccozyma* (Botha, 2011), most species have a fragmented geographical distribution with few shared species between locations. One study found that only eight of the 57 species found in soils of Mediterranean xerophyl forests were shared between the three sampling plots in the same locality (Yurkov et al., 2016a). Another study found only a single species to be present in all three sampled temperate forests in Germany (Yurkov et al., 2012). While these results highlight the need for continued and expanded environmental surveillance to characterize soil yeasts and identify novel species, current sampling efforts are mostly concentrated in the developed world with a distinct lack of environmental sampling in some Asian and African regions. Furthermore, most environmental surveys reported in the literature are geographically constrained, with sampling often limited to one type of soil within a single locality, region, or country (Into et al. 2020; A. H. Li et al. 2020; Monteiro Moreira and Martins do Vale 2020; Tepeeva, Glushakova, and Kachalkin 2018).

Inadequate sampling, combined with a lack of large-scale yeast-focused studies, has left a significant portion of the global soil yeast diversity untapped. In 2014, Tedersoo and colleagues used metagenomics to conduct a global investigation of soil fungal diversity encompassing 39 countries, including several under-sampled regions in Asian and African continents (Tedersoo et al. 2014). In this culture-independent study, they performed high-throughput sequencing of the ITS region on DNA directly extracted from soil samples to characterize soil fungal communities of sampled locations. Their analyses identified mean annual precipitation and distance from equator as the two most significant predictors of soil fungal diversity worldwide. In 2019, Egidi and colleagues also relied on culture-independent, high-throughput sequencing of the ITS region to characterize soil fungal communities in 235 soil samples collected from 18 countries across the globe (Egidi et al. 2019). Consistent with the previous study, their findings implicated vegetation and climatic factors as the most deterministic of soil fungal community dynamics. While both studies revealed global patterns and predictors of soil fungal diversity overall, neither focused on yeasts as a group of interest, leading to limited insights on global soil yeast distributions. Given the genetic and phylogenetic diversity among yeast species,

culture-dependent methods are more suitable for isolating yeasts from soil and characterizing soil yeast communities.

1.3 Human pathogenic yeast *Cryptococcus deneoformans*

Cryptococcus deneoformans is a basidiomycetous, encapsulated yeast commonly found in soil as well as in association with pigeon excreta and tree barks. Along with the sister species *C. neoformans,* it is the causative agent of cryptococcosis, a term given for a range of opportunistic infections that include skin infections, respiratory infections, and fungal meningitis (Charlier et al. 2009; Goldman, Lee, and Casadevall 1994; Husain, Wagener, and Singh 2001) . 220 000 global cases of cryptococcosis are reported every year with 180 000 resulting in death (Rajasingham et al. 2017). The genomes of the two sister species show 85-90% sequence similarity with the two species believed to have diverged from the last common ancestor 20 million years ago (Kavanaugh, Fraser, and Dietrich 2006; O'Hanlon et al. 2018). The first scientific record of *C. deneoformans* dates back to 1894 when Otto Busse and Abraham Busch isolated a '*Saccharomyces-like'* organism from a patient's bone infection (Freij and Freij 2015). In 1901, Jean-Paul Vuillemin noticed this yeast to lack some of the defining characteristics of the *Saccharomyces* genus, specifically the lack of ascospore production, and thus officially renamed it *Cryptococcus neoformans* (Srikanta, Santiago-Tirado, and Doering 2014). In the 1950s, a series of studies showed that *C. neoformans* strains can be categorized into four distinct serotypes (A, B, C and D) based on unique antigenic and morphological properties (Evans 1950; Wilson, Bennett, and Bailey 1968). Serotype D was eventually elevated to variety status and named *Cryptococcus neoformans* var. *neoformans,* until in 2015, it was recognized as a distinct species named *Cryptococcus deneoformans* (Hagen et al. 2015; Kwon-Chung and Varma 2006).

C. deneoformans is commonly isolated from environmental and clinical sources in temperate climates, most notably in Europe where it is responsible for over 30% of all reported cryptococcal infections (Bovers, Hagen, & Boekhout, 2008; Kwon-Chung & Bennett, 1984a). The infectious agent of *C. deneoformans* is thought to be the basidiospores which germinate and establish infections in the lungs upon inhalation (Velagapudi et al. 2009). The yeast cells can then potentially enter the bloodstream

and disseminate throughout the rest of the body: if it crosses the blood-brain barrier of the central nervous system, meningoencephalitis can arise (Chang et al. 2004). *C. deneoformans* has three essential virulence factors, namely thermotolerance (i.e., ability to grow at >37°C), production of a protective polysaccharide capsule around the cell, and the production of melanin pigments. All three play crucial roles in counteracting antimicrobial activities of the host's immune system (Kozel 1995). These 'dual-role' phenotypes confer resistance to common environmental stressors such as heat, UV light and oxygen/nitrogen radicals in its natural habitat while also facilitating the transition of *C. deneoformans* into an opportunistic pathogen upon entering vulnerable hosts (Casadevall, Steenbergen, and Nosanchuk 2003).

Given the beneficial role of melanin pigments in both environmental and clinical settings, it is interesting to observe significant intraspecific variation in pigmentation among *C. deneoformans* strains in natural populations (Figure 2). *C. deneoformans* oxidizes exogenous ortho-phenolic compounds such as L-dopamine (L-DOPA) to produce DOPA-melanin, a type of eumelanin closely related to melanin pigments found in human skin (Williamson, Wakamatsu, and Ito 1998). Melanin biosynthesis in *C. deneoformans* is catalyzed by a laccase enzyme: two laccase genes, *LAC1* and *LAC2*, are found in tandem on chromosome 7. *LAC1* enzyme is the primary catalyzer with mutations in *LAC2* gene only causing slight delays in melanin formation (Pukkila-Worley et al., 2005b). The basal transcript level of *LAC2* was also found to be significantly lower than that of *LAC1* (Pukkila-Worley et al., 2005b). Previous work from our group showed that nucleotide variations in *LAC1* are significantly correlated with melanin variance among clinical and environmental strains of *C. deneoformans* (Samarasinghe et al. 2018; Vogan et al. 2016). A recent quantitative trait locus (QTL) analysis identified *RIC8*, which encodes a regulator of cAMP-PKA signaling, as another contributor to melanin variance (Roth et al. 2021). While not directly involved in the catalysis of melanin synthesis, many genes have been implicated in the direct regulation of *LAC1* and thus melanin production in *C. deneoformans.* For an example, four transcription factors, Bzp4, Usv101, Mbs1, and Hob1, are required for the induction of *LAC1* expression (Lee et al. 2019). Whether

polymorphisms at these genetic loci contribute to melanin variance observed among natural strains remains to be investigated.

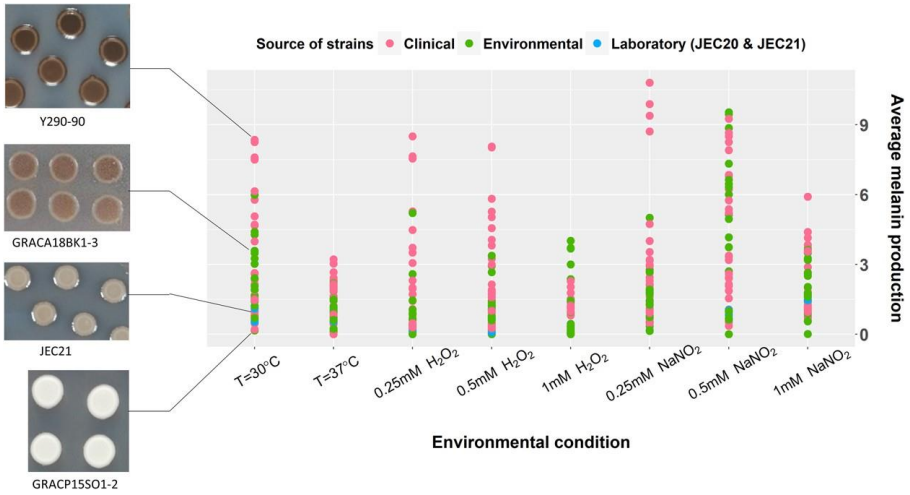


Figure 2: Variation in melanin production observed among natural strains of *C. deneoformans*. Melanin production of 54 global strains was quantified in eight different environmental conditions. From left to right, these are 30°C (optimum growth temperature for *C. deneoformans*), 37°C (thermal stress), low, intermediate and high oxidative stress, and low, intermediate and high nitrosative stress. In most conditions, the strains showed significant variation in melanin production. Figure adapted from Figure 1 in Samarasinghe et al. (2018).

1.4 Hybrids of *Cryptococcus deneoformans* and *Cryptococcus neoformans*

The closely related, haploid sister species, *C. neoformans* (serotype A) and *C. deneoformans* (serotype D) are known to hybridize in nature, giving rise to diploid/aneuploid, hybrid offspring commonly referred to as AD hybrids (Samarasinghe and Xu 2018). The first report of AD hybrids came in 1977 where an epidemiological survey from the United States found 11 of the 272 tested cryptococcal isolates to consistently react with both A and D antisera and were therefore typed as AD hybrids (Bennett et al., 1977). Interestingly, a strain isolated from peach juice in 1894, originally thought to be *C. deneoformans,* was recently determined to be an AD hybrid (Boekhout et al., 2001; Sanfelice, 1894; Viviani et al., 1997). AD hybrids are becoming increasing prevalent in clinical settings, especially in Mediterranean Europe where up to 40% of cryptococcal infections are now caused by hybrid strains (Cogliati et al. 2001). Recent global dispersal of *C. neoformans* and *C. deneoformans,* which were previously geographically isolated to tropical and temperate climates

respectively, is likely contributing to the ongoing hybridization between the two species (Xu, Vilgalys, and Mitchell 2000).

Hybrid spores generated in the lab via mating between *C. neoformans* and *C. deneoformans* strains have low viability with approximately 5% successfully germinating to yield AD hybrid strains (Lengeler, Cox, and Heitman 2001). Spores that do germinate often contain an abnormal number of chromosomes, indicating impaired meiosis and nondisjunction of parental chromosomes during sexual sporulation (Lengeler, Cox, and Heitman 2001; Litvintseva et al. 2007; Sun and Xu 2007; Tanaka, Nishimura, and Miyaji 1999). Hybrid genomes are highly plastic with loss of heterozygosity (LOH) frequently observed during vegetative growth via chromosomal loss and/or mitotic gene conversion (Hu et al. 2008; W. Li et al. 2012). Both lab-derived and natural AD hybrids show remarkable phenotypic diversity in both virulence and virulence-associated phenotypes. While some AD hybrids are similar in virulence to their more virulent parent, others show significant hybrid vigour and surpass both parents in virulence (Aminnejad et al. 2016; M. Cogliati et al. 2011; Vogan et al. 2016). A few studies found AD hybrids to exceed both parents in their resistance to UV irradiation and the antifungal drug FK506 (Chaturvedi et al. 2002; M. Li et al. 2012; W. Li et al. 2012; Lin et al. 2007; Litvintseva et al. 2007). Another study found extreme phenotypes, both positive (i.e. significantly superior to both parents) and negative (i.e. significantly inferior to both parents), among lab-derived AD hybrids in the expression of virulence factors including capsule production, growth at 37°C and melanin production (Vogan et al. 2016).

AD hybrids' remarkable phenotypic diversity is likely a product of their dynamic and unique genetic make-up. AD hybrids can be an ideal model system for studying outcomes of hybridization among fungal lineages. Patterns of parental allele distribution across AD hybrid genomes can reveal the nature of interactions that could occur when two independently evolving, closely related genomes are brought into contact through human interference.

1.5 Objectives

Yeast species have been closely linked to many aspects of human health and wellbeing for centuries. My PhD thesis aimed to expand our current knowledge of global diversity and distribution patterns of culturable soil yeasts, and the genetics of intraspecific melanin variance and hybridization in the human pathogenic soil yeast, *C. deneoformans*. My thesis had six main objectives which are described in detail through chapters two to seven. In chapter 2, I investigated the global distribution of soil yeasts by characterizing 1473 yeast strains isolated from soils of nine countries in six continents. In chapter 3, I investigated the population structure and molecular epidemiology of a novel *C. deneoformans* population we discovered in desert soils of Saudi Arabia. In chapter 4, I conducted a bulk segregant analysis using Illumina whole genome sequencing data to identify genetic loci contributing to variation in melanin pigmentation among *C. deneoformans* strains. In chapter 5, I identified AD hybrids as a model system for studying hybridization between fungal lineages by synthesizing the literature on genetics and virulence of cryptococcal hybrids. In chapter 6, I summarized the current state of knowledge on cryptococcal hybrids and made recommendations for future research. In chapter 7, I investigated patterns of parental allele distribution in a lab-derived AD hybrid population to determine the interactions of the two parents' chromosomes within hybrid genomes.

1.6 References

Aminnejad, Mojgan et al. 2016. "Identification and Characterization of VNI/VNII and Novel VNII/VNIV Hybrids and Impact of Hybridization on Virulence and Antifungal Susceptibility Within the C. Neoformans/C. Gattii Species Complex." *PLOS ONE* 11(10): e0163955. http://dx.plos.org/10.1371/journal.pone.0163955.

Arora, Parth et al. 2021. "Environmental Isolation of Candida Auris from the Coastal Wetlands of Andaman Islands, India." *mBio* 12(2): 1–9.

Barnett, James A. 2004. "A History of Research on Yeasts 8: Taxonomy." *Yeast* 21(14): 1141–93. https://onlinelibrary.wiley.com/doi/full/10.1002/yea.1154 (August 4, 2021).

Blackwell, Meredith. 2011. "The Fungi: 1, 2, 3 ... 5.1 Million Species?" *American journal of botany* 98(3): 426–38. http://www.ncbi.nlm.nih.gov/pubmed/21613136 (June 30, 2017).

Bongomin, Felix, Sara Gago, Rita O. Oladele, and David W. Denning. 2017. "Global and Multi-National Prevalence of Fungal Diseases—Estimate Precision." *Journal of Fungi* 3(4). /pmc/articles/PMC5753159/ (August 2, 2021).

Botha, Alfred. 2011. "The Importance and Ecology of Yeasts in Soil." *Soil Biology and Biochemistry* 43(1): 1–8.

Casadevall, Arturo, Judith N Steenbergen, and Joshua D Nosanchuk. 2003. "'Ready Made' Virulence and 'Dual Use' Virulence Factors in Pathogenic Environmental Fungi - the Cryptococcus Neoformans Paradigm." *Current Opinion in Microbiology* 6(4): 332–37. http://linkinghub.elsevier.com/retrieve/pii/S1369527403000821 (June 28, 2017).

Chang, Yun C. et al. 2004. "Cryptococcal Yeast Cells Invade the Central Nervous System via Transcellular Penetration of the Blood-Brain Barrier." *Infection and Immunity* 72(9): 4985–95. https://journals.asm.org/journal/iai (August 4, 2021).

Charlier, Caroline et al. 2009. "Evidence of a Role for Monocytes in Dissemination and Brain Invasion by Cryptococcus Neoformans." *Infection and Immunity*

77(1): 120–27.

Chaturvedi, Vishnu et al. 2002. "Molecular Genetic Analyses of Mating Pheromones Reveal Intervariety Mating or Hybridization in Cryptococcus Neoformans." *Infection and immunity* 70(9): 5225–35. http://www.ncbi.nlm.nih.gov/pubmed/12183574 (October 16, 2017).

Choi, JaeJin, and Sung-Hou Kim. 2017. "A Genome Tree of Life for the Fungi Kingdom." *Proceedings of the National Academy of Sciences* 114(35): 9391–96. https://www.pnas.org/content/114/35/9391 (August 4, 2021).

Cogliati, M., F. Barchiesi, E. Spreghini, and Anna Maria Tortorano. 2011. "Heterozygosis and Pathogenicity of Cryptococcus Neoformans AD-Hybrid Isolates."*Mycopathologia*.173(5–6):347–57. http://link.springer.com/10.1007/s11046-011-9467-x (September 12, 2018).

Cogliati, M et al. 2001. "Origin of Cryptococcus Neoformans Var. Neoformans Diploid Strains." *Journal of clinical microbiology* 39(11): 3889–94. http://www.ncbi.nlm.nih.gov/pubmed/11682503 (December 18, 2017).

Cogliati, Massimo. 2013. "Global Molecular Epidemiology of Cryptococcus Neoformans and Cryptococcus Gattii: An Atlas of the Molecular Types." *Scientifica* 2013. http://dx.doi.org/10.1155/2013/675213.

Connell, L. et al. 2008. "Diversity of Soil Yeasts Isolated from South Victoria Land, Antarctica." *Microbial Ecology* 56(3): 448–59. https://link.springer.com/article/10.1007/s00248-008-9363-1 (February 27, 2021).

Egidi, Eleonora et al. 2019. "A Few Ascomycota Taxa Dominate Soil Fungal Communities Worldwide." *Nature Communications* 10(1): 1–9. https://doi.org/10.1038/s41467-019-10373-z (March 22, 2021).

Evans, E. E. 1950. "The Antigenic Composition of Cryptococcus Neoformans. I. A Serologic Classification by Means of the Capsular and Agglutination Reactions." *Journal of Immunology* 64(5): 423–30.

Freij, Joudeh B., and Bishara J. Freij. 2015. "The Earliest Account of Human Cryptococcosis (Busse-Buschke Disease) in a Woman with Chronic Osteomyelitis of the Tibia." *Pediatric Infectious Disease Journal* 34(11): 1278. https://journals.lww.com/pidj/Fulltext/2015/11000/The_Earliest_Account_of_Human_Cryptococcosis.33.aspx (August 4, 2021).

Goldman, D., S. C. Lee, and A. Casadevall. 1994. "Pathogenesis of Pulmonary Cryptococcus Neoformans Infection in the Rat." *Infection and Immunity* 62(11): 4755–61.

Hagen, Ferry et al. 2015. "Recognition of Seven Species in the Cryptococcus Gattii/Cryptococcus Neoformans Species Complex." *Fungal Genetics and Biology* 78: 16–48.

Hu, Guanggan et al. 2008. "Comparative Hybridization Reveals Extensive Genome Variation in the AIDS-Associated Pathogen Cryptococcus Neoformans." *Genome biology* 9(2): R41. http://www.ncbi.nlm.nih.gov/pubmed/18294377 (March 4, 2018).

Husain, S., M. M. Wagener, and N. Singh. 2001. "Cryptococcus Neoformans Infection in Organ Transplant Recipients: Variables Influencing Clinical Characteristics and Outcome." *Emerging infectious diseases* 7(3): 375–81.

Into, Parichat, Ana Pontes, José Paulo Sampaio, and Savitree Limtong. 2020. "Yeast Diversity Associated with the Phylloplane of Corn Plants Cultivated in Thailand." *Microorganisms* 8(1): 80. https://www.mdpi.com/2076-2607/8/1/80 (March 29, 2021).

Jeffery-Smith, Anna et al. 2018. "Candida Auris: A Review of the Literature." *Clinical Microbiology Reviews* 31(1). https://doi.org/10.1128/CMR.00029-17. (August 2, 2021).

Kavanaugh, Laura A., James A. Fraser, and Fred S. Dietrich. 2006. "Recent Evolution of the Human Pathogen Cryptococcus Neoformans by Intervarietal Transfer of a 14-Gene Fragment." *Molecular Biology and Evolution* 23(10): 1879–90. http://academic.oup.com/mbe/article/23/10/1879/1097209/Recent-Evolution-of-

the-Human-Pathogen (March 1, 2018).

Kozel, Thomas R. 1995. "Virulence Factors of Cryptococcus Neoformans." *Trends in Microbiology* 3(8): 295–99.

Kurtzman, Cletus, J.W. Fell, and Teun Boekhout, eds. 2011. *The Yeasts A Taxonomic Study*. 5th edition. Elsevier. https://www.elsevier.com/books/the-yeasts/kurtzman/978-0-444-52149-1 (February 24, 2021).

Kwon-Chung, Kyung J., and Ashok Varma. 2006. "Do Major Species Concepts Support One, Two or More Species within *Cryptococcus Neoformans*?" *FEMS Yeast Research* 6(4): 574–87. https://academic.oup.com/femsyr/article-lookup/doi/10.1111/j.1567-1364.2006.00088.x (September 8, 2018).

Lee, Dongpil et al. 2019. "Unraveling Melanin Biosynthesis and Signaling Networks in Cryptococcus Neoformans." *mBio* 10(5).

Lengeler, K B, G M Cox, and J Heitman. 2001. "Serotype AD Strains of Cryptococcus Neoformans Are Diploid or Aneuploid and Are Heterozygous at the Mating-Type Locus." *Infection and immunity* 69(1): 115–22. http://www.ncbi.nlm.nih.gov/pubmed/11119496 (October 16, 2017).

Li, A. H. et al. 2020. "Diversity and Phylogeny of Basidiomycetous Yeasts from Plant Leaves and Soil: Proposal of Two New Orders, Three New Families, Eight New Genera and One Hundred and Seven New Species." *Studies in Mycology* 96: 17–140.

Li, Meng et al. 2012. "Antifungal Susceptibilities of Cryptococcus Species Complex Isolates from AIDS and Non-AIDS Patients in Southeast China." *Brazilian Journal of Infectious Diseases*.

Li, Wenjun et al. 2012. "Genetic Diversity and Genomic Plasticity of Cryptococcus Neoformans AD Hybrid Strains." *G3 (Bethesda, Md.)* 2(1): 83–97. http://www.ncbi.nlm.nih.gov/pubmed/22384385 (December 12, 2017).

Lin, Xiaorong et al. 2007. "AADα Hybrids of Cryptococcus Neoformans: Evidence of Same-Sex Mating in Nature and Hybrid Fitness." *PLoS Genetics* 3(10): e186.

http://dx.plos.org/10.1371/journal.pgen.0030186 (October 10, 2017).

Litvintseva, Anastasia P. et al. 2007. "Many Globally Isolated AD Hybrid Strains of Cryptococcus Neoformans Originated in Africa." *PLoS Pathogens* 3(8): 1109–17. http://dx.plos.org/10.1371/journal.ppat.0030114 (October 22, 2017).

McGovern, Patrick E. et al. 2004. "Fermented Beverages of Pre- and Proto-Historic China." *Proceedings of the National Academy of Sciences* 101(51): 17593–98. https://www.pnas.org/content/101/51/17593 (August 2, 2021).

Monteiro Moreira, Geisianny Augusta, and Helson Mario Martins do Vale. 2020. "Soil Yeast Communities in Revegetated Post-Mining and Adjacent Native Areas in Central Brazil." *Microorganisms* 8(8): 1116. https://www.mdpi.com/2076-2607/8/8/1116 (March 29, 2021).

O'Hanlon, Simon J. et al. 2018. "Recent Asian Origin of Chytrid Fungi Causing Global Amphibian Declines." *Science (New York, N.Y.)* 360(6389): 621. /pmc/articles/PMC6311102/ (August 1, 2021).

Opulente, Dana A et al. 2019. "Pathogenic Budding Yeasts Isolated Outside of Clinical Settings." *FEMS Yeast Research* 19(3). https://academic.oup.com/femsyr/article/19/3/foz032/5479245 (August 2, 2021).

Papon, Nicolas, Vincent Courdavault, Marc Clastre, and Richard J. Bennett. 2013. "Emerging and Emerged Pathogenic Candida Species: Beyond the Candida Albicans Paradigm." *PLOS Pathogens* 9(9): e1003550. https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1003550 (August 2, 2021).

Rajasingham, Radha et al. 2017. "Global Burden of Disease of HIV-Associated Cryptococcal Meningitis: An Updated Analysis." *The Lancet Infectious Diseases* 17(8): 873–81.

Roth, Cullen et al. 2021. "Pleiotropy and Epistasis within and between Signaling Pathways Defines the Genetic Architecture of Fungal Virulence." *PLOS Genetics* 17(1):e1009313. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1009313

(July 6, 2021).

Samarasinghe, Himeshi et al. 2018. "Genetic Factors and Genotype-Environment Interactions Contribute to Variation in Melanin Production in the Fungal Pathogen Cryptococcus Neoformans." *Scientific Reports* 8(1): 9824. http://www.nature.com/articles/s41598-018-27813-3 (July 14, 2018).

Samarasinghe, Himeshi, and Jianping Xu. 2018. "Hybrids and Hybridization in the Cryptococcus Neoformans and Cryptococcus Gattii Species Complexes." *Infection, Genetics and Evolution* 66: 245–55. https://www.sciencedirect.com/science/article/pii/S1567134818302624 (July 3, 2019).

Schoch, Conrad L. et al. 2012. "Nuclear Ribosomal Internal Transcribed Spacer (ITS) Region as a Universal DNA Barcode Marker for Fungi." *Proceedings of the National Academy of Sciences of the United States of America* 109(16): 6241–46. www.pnas.org/cgi/doi/10.1073/pnas.1117018109 (July 4, 2021).

Sobel, J. D. et al. 1998. "Vulvovaginal Candidiasis: Epidemiologic, Diagnostic, and Therapeutic Considerations." *American Journal of Obstetrics and Gynecology* 178(2): 203–11.

Srikanta, Deepa., Felipe H. Santiago-Tirado, and Tamara L. Doering. 2014. "Cryptococcus Neoformans: Historical Curiosity to Modern Pathogen." *Yeast (Chichester, England)* 31(2): 47. /pmc/articles/PMC3938112/ (August 4, 2021).

Sun, Sheng, and Jianping Xu. 2007. "Genetic Analyses of a Hybrid Cross between Serotypes A and D Strains of the Human Pathogenic Fungus Cryptococcus Neoformans." *Genetics* 177(3): 1475–86. http://www.ncbi.nlm.nih.gov/pubmed/17947421 (October 16, 2017).

Tanaka, R, K Nishimura, and M Miyaji. 1999. "Ploidy of Serotype AD Strains of Cryptococcus Neoformans." *Japanese journal of medical mycology* 40(1): 31–34. http://www.ncbi.nlm.nih.gov/pubmed/9929580 (February 28, 2018).

Tedersoo, Leho et al. 2014. "Global Diversity and Geography of Soil Fungi." *Science* 346(6213). http://science.sciencemag.org/content/346/6213/1256688 (July 2,

2017).

Tepeeva, A. N., A. M. Glushakova, and A. V. Kachalkin. 2018. "Yeast Communities of the Moscow City Soils." *Microbiology (Russian Federation)* 87(3): 407–15. www.mycobank.org (March 29, 2021).

Velagapudi, Rajesh et al. 2009. "Spores as Infectious Propagules of Cryptococcus Neoformans." *Infection and immunity* 77(10): 4345–55. http://www.ncbi.nlm.nih.gov/pubmed/19620339 (March 28, 2018).

Vishniac, H. S. 1996. "Biodiversity of Yeasts and Filamentous Microfungi in Terrestrial Antarctic Ecosystems." *Biodiversity and Conservation* 5(11): 1365–78. https://link.springer.com/article/10.1007/BF00051983 (February 27, 2021).

Vogan, Aaron A., Jordan Khankhet, Himeshi Samarasinghe, and Jianping Xu. 2016. "Identification of QTLs Associated with Virulence Related Traits and Drug Resistance in Cryptococcus Neoformans." *G3 (Bethesda, Md.)* 6(9): 2745–2759.

Williamson, P. R., K. Wakamatsu, and S. Ito. 1998. "Melanin Biosynthesis in Cryptococcus Neoformans." *Journal of Bacteriology* 180: 1570–72.

Wilson, D.E., J.E. Bennett, and J.W. Bailey. 1968. "Serological Grouping of Cryptococcus Neoformans." *Society for Experimental Biology and Medicine* 127: 820–23.

Wu, Bing et al. 2019. "Current Insights into Fungal Species Diversity and Perspective on Naming the Environmental DNA Sequences of Fungi." *http://mc.manuscriptcentral.com/tmyc* 10(3): 127–40. https://www.tandfonline.com/doi/abs/10.1080/21501203.2019.1614106 (July 19, 2021).

Xu, Jianping. 2016. "Fungal DNA Barcoding." *Genome* 59(11): 913–32. http://www.ncbi.nlm.nih.gov/Taxonomy/ (March 30, 2021).

Xu, Jianping, Rytas Vilgalys, and Thomas G. Mitchell. 2000. "Multiple Gene Genealogies Reveal Recent Dispersion and Hybridization in the Human Pathogenic Fungus *Cryptococcus Neoformans*." *Molecular Ecology* 9(10): 1471–

81. http://doi.wiley.com/10.1046/j.1365-294x.2000.01021.x (June 17, 2016).

Yang, Yun-Liang et al. 2012. "Comparison of Human and Soil Candida Tropicalis Isolates with Reduced Susceptibility to Fluconazole." *PLoS ONE* 7(4). /pmc/articles/PMC3320620/ (August 2, 2021).

Yurkov, Andrey M. 2018. "Yeasts of the Soil - Obscure but Precious." *Yeast* 35(5): 369–78. http://doi.wiley.com/10.1002/yea.3310 (February 3, 2021).

Chapter 2

Global Patterns in Culturable Soil Yeast Diversity

2.1 Preface

Soil was not discovered to be a primary habitat and reservoir of yeasts until the 1950s. Since then, significant strides have been made in characterizing functions, metabolic profiles, and interactions of soil yeast communities. Lack of environmental sampling, especially in Asian and African regions, has led to global soil yeast diversity being significantly underestimated. Most previous studies investigating soil yeasts were geographically constrained, with sampling limited to a single country, region, or city. Here, we used thousands of soil samples to characterize the culturable soil yeast diversity in nine countries across six continents. Our study found global yeast diversity to be significantly correlated with climatic factors as well as anthropogenic travel.

This study is now published in iScience volume 24, issue 10, article number 103098. I am the primary contributor of this work. I conducted the majority of the experiments, as well as the analyses and drafting of the manuscript. Co-authors of this article are Yi Lu, Renad Aljohani, Ahmad Al-Amad, Heather Yoell and Jianping Xu.

2.2 Abstract

Yeasts, broadly defined as unicellular fungi, fulfill essential roles in soil ecosystems as decomposers and nutrition sources for fellow soil-dwellers. Broad-scale investigations of soil yeasts pose a methodological challenge as metagenomics are of limited use for identifying this group of fungi. Here we characterize global soil yeast diversity using fungal DNA barcoding on 1473 yeasts cultured from 3826 soil samples obtained from nine countries in six continents. We identify mean annual precipitation and international air travel as two significant correlates with soil yeast community structure and composition worldwide. Evidence for anthropogenic influences on soil yeast communities, directly via travel and indirectly via altered rainfall patterns resulting from climate change, is concerning as we found common infectious yeasts frequently distributed in soil in several countries. Our discovery of 41 putative novel species highlights the continued need for culture-based studies to advance our knowledge of environmental yeast diversity.

2.3 Introduction

Soil is host to an incredible amount of microbial life, with each gram containing over 10 billion cells of bacteria, archaea and fungi (Roesch et al., 2007). Despite their relatively low abundance, soil fungi fulfill essential roles in decomposition of organic material, nutrient cycling and soil fertilization (Frac et al., 2018). This is especially true for yeasts, broadly defined as unicellular fungi, whose numbers rarely exceed thousands of cells per gram of soil. Yet, yeasts in soil ecosystems are essential decomposers and nutrient sources for fellow soil-dwelling protists, bacteria, insects, and nematodes (Botha, 2011; Yurkov, 2018). In fact, yeasts may be the predominant soil fungi in cold biospheres such as continental Antarctica (Connell et al., 2008; Vishniac, 1996). Soil is also a primary reservoir for some *Candida* and *Cryptococcus* species that are opportunistic pathogens of humans (Kurtzman et al., 2011).

It is becoming increasingly apparent that the true extent of global soil yeast diversity is significantly underestimated. While yeast cells were first observed under the microscope in 1680 by Anton Van Leeuwenhoek, their natural habitats were a topic of contention among mycologists who often associated yeasts with fruit trees and

fermentation. It was not until the 1950s that soil was established as a true natural habitat of yeasts where they live and reproduce. The culturing media, incubation temperatures and techniques used in pioneering studies were expanded in later projects to isolate soil yeasts of diverse metabolic and functional profiles (reviewed in Yurkov, 2018). Currently, the diverse array of yeasts recovered and characterized from soils across the globe contribute to the ~1500 recognized yeast species on the planet (Kurtzman et al., 2011; Naranjo-Ortiz and Gabaldón, 2019). The yeast discovery rate has accelerated since the turn of this century, with almost 50% of the known yeast species described since 2011 (Wu et al., 2019; Yurkov, 2018; Naranjo-Ortiz and Gabaldón, 2019). Environmental surveys routinely uncover novel yeast species, accounting for as much as 30% of yeast populations, highlighting the need to revise current estimates of global yeast diversity (Groenewald et al., 2018; Yurkov et al., 2016b, 2016a).

Lack of adequate environmental sampling, especially in Asia, Africa, South America and Central America, limits the discovery of novel yeast species, characterization of soil yeast communities, and prediction of global yeast diversity patterns. Soil yeast populations often differ in structure and composition between locations. With the exception of a few genera that are widespread in soil such as *Cyberlindnera, Schwanniomyces, Naganishia, Goffeauzyma* and *Solic ο ccozyma* (Botha, 2011), most yeast species have a fragmented distribution with a few shared species between sites, even within the same geographical region. One study found that only eight of the 57 species found in soils of Mediterranean xerophyl forests were shared between the three sampling plots in the same locality (Yurkov et al., 2016a). Another study found only a single species to be present in all three sampled temperate forests in Germany (Yurkov et al., 2012). So far, most environmental surveys reported in the literature have been ecologically/ geographically limited, with sampling often focusing on a specific ecological niche within a single locality, region, or country (Into et al., 2020; Li et al., 2020; Monteiro Moreira and Martins do Vale, 2020; Tepeeva et al., 2018).

Factors affecting soil yeast diversity have not been fully elucidated but soil moisture, soil pH, carbon content and nitrogen content have been implicated as contributing variables (reviewed in Botha, 2011). In 2006, Vishniac analyzed prominent yeast

species in soil along a latitudinal gradient and found that variations in mean annual temperature, mean annual rainfall and soil electrical conductivity explained ~44% of the total variance in yeast species distributions (Vishniac, 2006). As their sampling locations were limited to the Americas and Antarctica, it is unknown whether the same trends persist on a global scale. One potential contributing factor to yeast species and genotype distributions is anthropogenic influences such as international travel. International travel has increased exponentially in the last few decades with direct implications for the global spread of organisms, most notably infectious disease agents and invasive species. The role of global travel in introducing infectious diseases to new areas and facilitating epidemics is well documented (Findlater and Bogoch, 2018), with the current COVID-19 pandemic being a prime example. International travel is likely affecting soil yeast communities by transferring previously geographically isolated species and genotypes across borders, although a link between the two has not been previously investigated.

Metagenomics is widely applied in the study of environmental microbes to investigate taxonomic diversity, characterize functional groups, and elucidate broad scale patterns (Abbasian et al., 2016; Abia et al., 2018; Egidi et al., 2019; Li and Qin, 2005). However, metagenomics and other culture-independent methods cannot be readily applied to the study of yeasts due to the lack of a suitable yeast-specific barcoding gene (Xu, 2016). This is mainly because yeasts are phylogenetically diverse and occur among filamentous fungi in two major phyla, Ascomycota and Basidiomycota, within the fungal kingdom. In a 2014 study that has not been surpassed in scale before or since, Tedersoo and colleagues used high throughput sequencing of fungal barcoding DNA to assess global soil fungal diversity and identify predictors of global diversity patterns (Tedersoo et al., 2014). Due to the lack of a sequence-based signature, yeasts were not singled out as a group of interest and thus limited information was presented on soil yeast diversity of the 39 countries included in the study. They identified mean annual precipitation and distance from equator as the two strongest overall predictors of soil fungal diversity on a global scale. It is not clear if and to what extent the same predictors apply to yeast diversity in soil.

Using a global collection of 3826 soil samples, here we assessed the culturable soil yeast diversity in nine countries representing all continents except Antarctica. We found soil yeast populations to be unique between countries in structure and composition, with 73% of the discovered yeast species found in only one of the sampled countries. Similar to that reported by Tedersoo et al. (2014), mean annual precipitation was the most significant predictor of culturable soil yeast diversity on a global scale. Importantly, we found air traffic volume to be significantly correlated with the number of shared species between countries, suggesting a potential link between international travel and transfer of yeast species across borders. Our study overcomes the geographical constraints of many previous studies by identifying soil yeast diversity patterns on a global scale. We also demonstrate that culture-dependent methods provide a more comprehensive framework than metagenomics for studying phylogenetically diverse, but morphologically targeted groups of organisms such as yeasts.

2.4 Materials and Methods

Soil collection

We collected soil from 53 locations in nine countries encompassing all continents except Antarctica (Figure 1 and Supplementary Table 1). At each location, we set up multiple plots: within each plot, we collected ten samples of topsoil (approximately 1g each) between 1-3 inches from the surface, using sterile protocols. Within each plot, the ten soil samples were at least 2m from each other. The 10 soil samples from the same plot were stored in the same sterile, 3cm x 7cm, resealable polyethylene bag to minimize bag usage and maximize soil representation from each plot. Once transported to our lab at McMaster University, Canada, we segregated the soil in each bag into ten 1g aliquots in 1.5ml Eppendorf tubes and stored at 4°C. Some bags had more than 10g of soil and those were separated into additional tubes. In total, this study investigated 3826 aliquot soil samples (from 380 bags) collected from the following countries: Cameroon (493 from 49 bags), Canada (300 from 30 bags), China (340 from 34 bags), Costa Rica (388 from 38 bags), France (327 from 37 bags), Iceland (316 from 31 bags), New Zealand (610 from 61 bags), Peru (490 from 49 bags) and Saudi Arabia (562 from 56 bags).

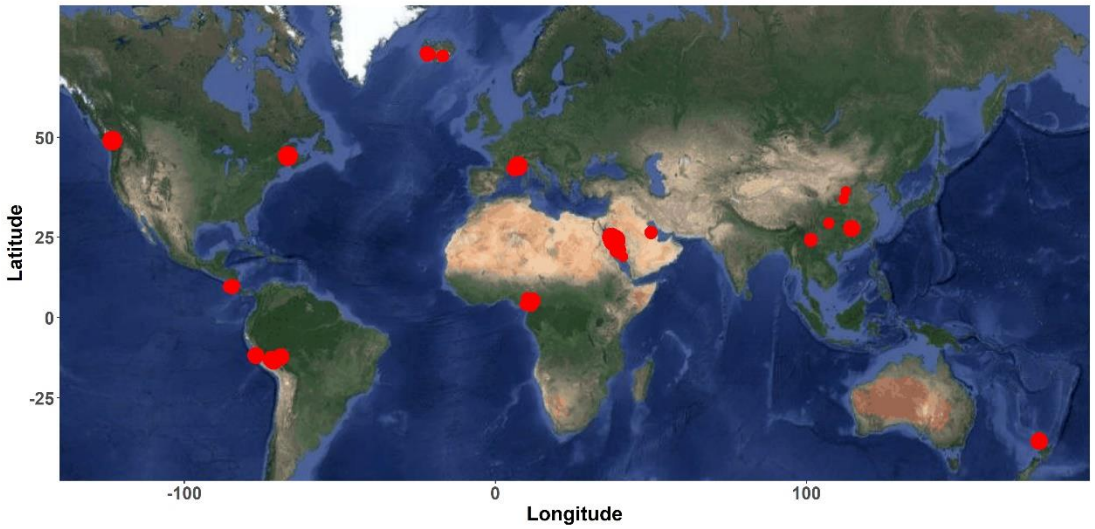


Figure 1: Soil sampling locations. 3826 soil samples were obtained from 53 locations, indicated by the red circles, in nine countries. The size of the circle corresponds to the number of samples obtained from that location.

Yeast isolation from soil samples

Yeasts were isolated at a temperature deemed to be optimal based on the country of origin's mean annual temperature. For each soil sample, approximately 0.1g was added into 5ml of YEPD broth (Yeast Extract-Peptone-Dextrose) in 13ml culture tubes and incubated in a roller drum for 24 hours. The broth contained the antibiotic chloramphenicol (50mg/L) and the selectively toxic fungicide benomyl (5mg/L) to inhibit bacterial and mold growth, respectively. For Iceland soil samples, we extended this incubation step to 72 hours due to slower yeast growth at 14°C. We then plated 100ul of the broth onto solid YEPD containing chloramphenicol and benomyl and incubated at the same temperature for an additional 2-5 days until microbial growth was visible. For each plate that contained morphologically yeast-like colonies, we randomly selected a representative colony and streaked it onto fresh YEPD plates to obtain single colonies. If more than one morphology was present, one representative colony of each type was separately streaked for single colonies. After 2-3 days' incubation, we randomly picked one single colony per yeast isolate and suspended in 50ul nuclease-free water to be used in Polymerase Chain Reaction (PCR).

Yeast identification via ITS sequencing

We identified the yeasts by sequencing their fungal barcoding gene, the ribosomal Internal Transcribed Spacer (ITS) regions. We performed colony PCR using primers ITS1 (5' TCCGTAGGTGAACCTGCGG 3') and ITS4 (5' TCCTCCGCTTATTGATATGC 3') to amplify the ITS region. The PCR cocktail consisted of 10ul Promega GoTaq Green Mastermix, 5ul nuclease-free water, 2ul each of the primers (2uM) and 1ul cell suspension. The thermocycling conditions were an initial denaturation step at 95°C for 10 minutes followed by 35 cycles of: (i) 95°C for 30 seconds, (ii) 55°C for 30 seconds, and (iii) 72°C for 1 minute. We ran 4ul of the PCR products on a 1.5% agarose gel to check for successful amplification. The remaining PCR products were sequenced using the Sanger method at Eurofins Genomics in Louisville, Kentucky (https://eurofinsgenomics.com/en/home/). We trimmed the low-quality ends of the ABI chromatograms generated from Sanger sequencing and batch converted to FASTA format with DNA Baser's ABI to FASTA converter software (www.DnaBaser.com). We used BLAST+ applications on the command line to query the multi-FASTA file against the NCBI nucleotide database to detect sequence similarity to existing ITS sequences. The BLAST searches were run remotely (-remote flag) to avoid downloading the entire database onto our servers. The output was compiled into a CSV (Comma Separated Values) file containing the top 10 matches for each query sequence. We inspected the CSV file manually to check for quality and for any inconsistencies in species identity within the top ten matches. We assigned species identities to our ITS sequences at a sequence similarity threshold of 98.41% to existing sequences in databases. This threshold was previously determined to be optimum to distinguish yeast species at the ITS locus based on an analysis of 9000 fungal sequences (Vu et al., 2016). Sequences with no matches surpassing this threshold were considered putative novel species. We performed a massBLASTer analysis for the putative novel sequences in the curated UNITE database (https://unite.ut.ee/) to identify the genus they are most closely affiliated with and the associated species hypothesis (SH) codes (Pante et al., 2015).

Statistical analyses of population diversity

All statistical analyses were conducted in RStudio v.4.0.2 using a combination of base functions and packages including ggmap (Kahle and Wickham, 2013), ggplot2 (Wickham, 2016), and tidyverse (Wickham et al., 2019). We quantified the diversity of yeast populations at our sampling sites by calculating the Shannon diversity index using the package Vegan v.2.5-7 (Oksanen et al., 2020). We conducted rarefaction analyses using the iNEXT package (Hsieh et al., 2016) to determine if sufficient soil sampling was performed in each of the nine countries to accurately estimate their culturable soil yeast diversity.

Relationship between yeast diversity and climate and geographic factors

Within each country, soil collection sites differed in climatic and environmental conditions, with the exception of New Zealand where sampling was limited to the metropolitan region of Auckland. We assigned the 53 sampling sites to 47 distinct locations based on their geographical coordinates. Using geographical coordinates, we calculated mean annual precipitation and mean annual temperature by averaging monthly data over a 26-year period from 1991-2016, available on Climate Change Knowledge Portal (https://climateknowledgeportal.worldbank.org/). We calculated the elevation of the sampling sites and their distance from the equator using Google Maps. We calculated the Shannon diversity index of the yeast populations found at the 47 distinct locations. We constructed mixed models using the package lme4 v.1.1-26 (Bates et al., 2015) where precipitation, temperature, elevation, and distance to equator were set as fixed effects, country was fitted as a random effect and Shannon diversity Index was fitted as the dependent variable.

Air traffic data

We extracted data on the number of trips occurring between each country-country pair over a 5-year period from 2011-2016 from the Global Transnational Mobility Dataset (Recchi et al., 2019). This dataset is compiled based on a combination of tourism data and distance-adjusted air-traffic data. Next, we calculated the number of yeast species shared between each country-country combination. To assess the correlation between

the number of shared species and the volume of air traffic between countries, a linear model was fitted between the two variables.

Comparison to metagenomics study

We compared our findings to a previous study that used culture-independent methods to investigate global diversity of soil fungi. In 2014, Tedersoo and colleagues extracted DNA directly from soil samples of 39 countries and performed high throughput sequencing of the ITS2 region using primers ITS3 and ITS4, covering a portion of the DNA barcoding fragment we sequenced here (Tedersoo et al., 2014). Four countries overlapped between the two studies, namely, Cameroon, Canada, China, and New Zealand. We conducted BLAST searches of our ITS sequences against the entire metagenomics dataset to identify the potential distributions of our cultured yeast species in the larger global soil samples. First, we extracted the ITS2 region from all of our full ITS sequences using the ITSx software (Bengtsson-Palme et al., 2013). Next, we used the QIIME VSEARCH tool to cluster the ITS2 sequences at 98.41% sequence identity into operational taxonomic units (OTU) (Rognes et al., 2016). Finally, we used the blastn functionality from the BLAST+ toolset to query our OTUs against the full OTU dataset from the metagenomics study. The output was compiled into a CSV file containing the top 5 matches for each query OTU sequence. This CSV file was perused manually to identify significant sequence similarity between query and match sequences.

2.5 Results

Yeast isolation and species identification

We isolated a total of 1473 yeasts from 3826 soil samples (Table 1). The yeast isolation rate varied among countries, ranging from 17% in Saudi Arabia to 87% in Canada. The yeast isolation rate and species distribution data from Cameroon soils have been reported in a previous study (Aljohani et al. 2018). Overall, we observed a slightly negative correlation between the number of soil samples and the yeast isolation rate ($p < 0.05$), as countries with more soil samples did not necessarily yield more yeast isolates. No statistically significant correlation was observed between yeast isolation rate and either the length of time between sampling and yeast isolation

in the lab, or the physical distance between sampling locations and our lab where the soil yeasts were isolated ($p > 0.10$).

We successfully assigned species identity to 1367 isolates using the 98.41% sequence identity cut-off to homologous ITS sequences in NCBI and UNITE databases. These strains were categorized into 90 species belonging to 37 genera, with *Candida* being the most species-rich genus (n=19 species). Six *Candida* species belonged to clade *Lodderomyces*, two belonged to *Pichia* clade, while the clades *Barnettozyma*, *Candida glaebosa*, *Cyberlindnera*, *Kurtzmaniella*, *Ogataea*, *Wickerhamomyces/Candida*, and *Yamadazyma* were represented by one species each. The four remaining *Candida* species have not been assigned to a clade (Supplementary Table 1). With 60 ascomycetes and 30 basidiomycetes, both major yeast-containing phyla within the fungal kingdom were broadly represented. The 90 species belonged to six Classes, ten Orders and 18 Families. However, two genera, *Nadsonia* and *Holtermanniella*, do not currently have defined Family associations (*incertae sedis*). The remaining 106 yeast strains can be grouped into 44 OTUs at 98.41% nucleotide similarity. They represent potentially novel yeast species since no existing sequences with >98.41% sequence identity to these OTUs were found in the databases. Genbank accession numbers to the ITS sequences of our 1473 isolates are MG817572 to MG817630 and MW894661 to MW896112 (Supplementary dataset 1).

Our rarefaction analyses suggested that sufficient soil sampling was conducted in each country to accurately estimate the true diversity of culturable soil yeasts at the respective sampling sites. Projections for Shannon diversity index beyond the number of soil samples included in the study revealed that the diversity of our yeast populations approached saturation asymptote (Supplementary Figure 1). Additional sampling in these locations was not likely to have revealed higher yeast species diversity.

Table 1: Summary statistics of yeast isolation from global soil samples

Country	Soil samples	Yeast isolates	Known species / Novel species	Country-specific species	Ascomycete species (isolates) / Basidiomycete species (isolates)	Shannon diversity index
Cameroon	493	110	10/9	12	18 (106) / 1 (4)	2.17
Canada	300	261	34/12	25	37 (179) / 9 (82)	3.06
China	340	230	23/5	15	15 (100) / 13 (130)	2.54
Costa Rica	388	95	20/2	9	18 (88) / 4 (7)	2.21
France	327	175	12/2	3	11 (172) / 3 (3)	1.26
Iceland	316	211	11/0	4	5 (25) / 6 (186)	1.25
New Zealand	610	155	14/4	5	6 (15) / 10 (137)	2.05
Peru	490	139	30/9	20	28 (113) / 10 (26)	3.27
Saudi Arabia	562	97	8/1	5	3 (3) / 6 (94)	0.91
Total	**3826**	**1473**	**90/44**	**98**		

Diversity and abundance of culturable soil yeast populations

The abundance and diversity of soil yeast populations varied significantly between countries. Saudi Arabia ranked lowest among the nine countries in the number of unique species, where 562 soil samples yielded 97 yeast isolates belonging to 9 species, one of which was novel. On the other hand, we obtained 261 yeast isolates from 300 Canadian soil samples, encompassing 46 species, 12 of which were novel. The number of yeast isolates and distinct species found in the seven remaining countries ranged from 95-230 and 11-39 respectively (Table 1). The Shannon diversity index of the soil yeast populations ranged from 0.91 (Saudi Arabia) to 3.27 (Peru). Less diverse populations tended to be dominated by a single yeast species, most notably in France, Iceland, and Saudi Arabia where *Candida subhashii*, *Goffeauzyma gastrica* and *Cryptococcus deneoformans* predominated the soil yeasts respectively (Figure 2).

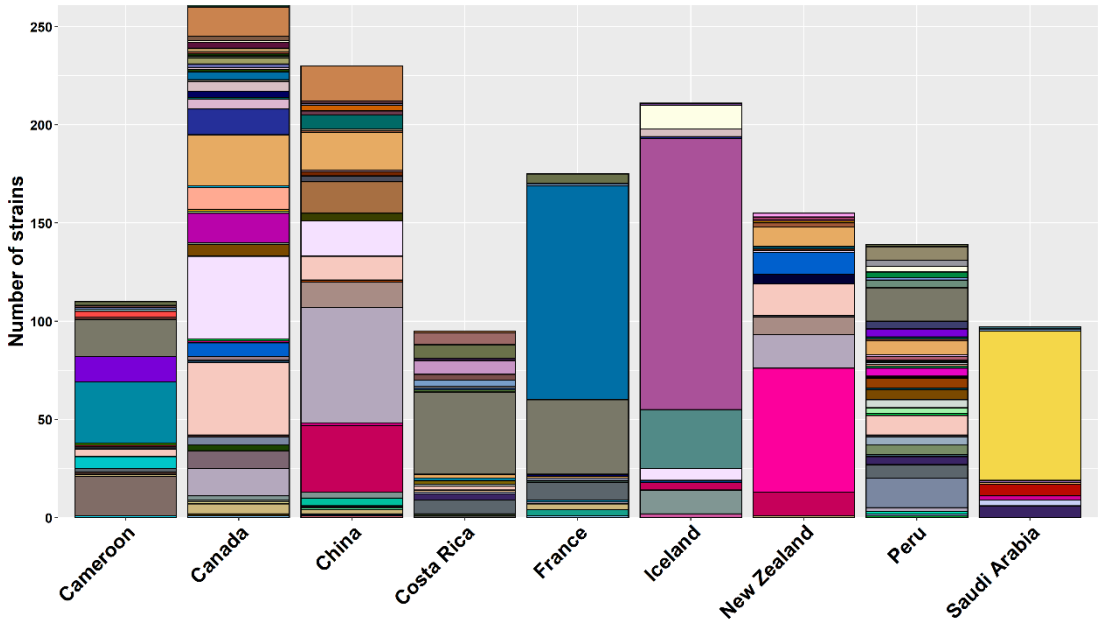


Figure 2: Culturable soil yeast populations by country. The X-axis represents the country. Each country is represented by a stacked bar plot. Each colour represents a unique species and the height of the colored sections indicate the abundance of that species.

Cameroon

As shown in the study by Aljohani et al. (2018), all but one of the 19 yeast species recovered from Cameroonian soil were ascomycetes. The population was dominated by four species: *Cyberlindnera subsufficiens* (28%), *Torulaspora globosa* (18%), *Candida tropicalis* (17%) and *Cyberlindnera saturnus* (12%). Overall, Cameroonian soil contained the second highest number of novel species (n=9) after Canada (n=12).

Canada

The culturable yeast population in Canadian soil was dominated by ascomycetes (37 species): the remaining 9 species were basidiomycetes. *Nadsonia starkeyi-henricii*, a little-known yeast that prefers relatively mild temperatures below 25°C, was the most abundant (16%) followed by the opportunistic pathogen *Papiliotrema laurentii* (14%). *Debaryomyces hansenii* (10%), *Barnettozyma californica* (6%) and *Candida* sp. (NEW) 8 (6%) were also present in significant amounts. Canadian soil contained 25 yeast species not found in soil samples of the other eight countries, including cold-adapted yeast *Cystofilobasidium ferigula*, industrial lactose fermenter *Kluyveromyces lactis*, and close relative of Baker's yeast, *Saccharomyces paradoxus.* Canada also had the highest number of novel species (n=12) of the nine sampled countries.

China

The Chinese culturable soil yeast population consisted of similar numbers of ascomycetes and basidiomycetes (15 and 13 respectively). This population was dominated by strains of *Solicoccozyma aeria* and *Solicoccozyma terrea* that together accounted for 40% of the population. Other yeasts with significant prevalence included *Debaryomyces hansenii* (8%), *Barnettozyma californica* (8%), *Nadsonia starkeyi-henricii* (8%) and *Candida* sp. (NEW) 6 (7%). 15 yeast species were only found in the Chinese soil which also yielded five novel species.

Costa Rica

Candida tropicalis was the dominant species in the culturable soil yeast population of Costa Rica with a prevalence of 44%. Frequencies of the remaining 21 species ranged from 1%-7%. Ascomycetous species outnumbered basidiomycetes at 18 to four. Costa Rica was notable for being the only sampled country to contain strains of the common opportunistic pathogens *Candida albicans* (6%) and *Candida orthopsilosis* (3%). Strains of pathogenic *Candida parapsilosis* were also present in Costa Rican soil (3%).

France

The majority of species in the French culturable soil yeast population were ascomycetes (n=11) while three species were basidiomycetes. *Candida subhashii*, an opportunistic pathogen first identified in 2009 (Adam et al., 2009), was the dominant yeast in this population with an abundance of 62%. *C. subhashii* was previously determined to have strong antagonistic activity against filamentous fungi and has potential as a biocontrol agent against plant pathogenic fungi (Hilber-Bodmer et al., 2017). The widespread opportunistic pathogen *Candida tropicalis* was the second most abundant species (22%), followed by *Saccharomyces cerevisiae* (5%). One strain of *Candida parapsilosis* was also detected in French soils.

Iceland

Of the 11 species isolated from Iceland soil, six were basidiomycetes and five were ascomycetes. With an abundance of 65%, *Goffeauzyma gastrica* was the dominant

species in the Iceland culturable soil yeast population. *G. gastrica* is a cold-tolerant yeast commonly isolated from environmental sources in Antarctica and is known for its production of antifreeze proteins (Białkowska et al., 2017; Ogaki et al., 2020; Villarreal et al., 2018). *Goffeauzyma gilvescens,* another cold-tolerant yeast commonly found in Antarctica, was the second most abundant (14%), followed by *Candida sake* (6%) and *Solicoccozyma terricola* (6%). Iceland was the only sampled country to not yield any novel yeast species.

New Zealand

Basidiomycete species (n=10) were slightly more prevalent than ascomycete species (n=8) in the New Zealand culturable soil yeast population. *Solicoccozyma phenolica* was the most abundant species with a prevalence of 41%, followed by *Solicoccozyma aeria* (11%), *Papiliotrema laurentii* (10%) and *Solicoccozyma terrea* (8%). We isolated several species with industrial potential from New Zealand soil including *Papiliotrema terrestris,* shown to produce β-galactosidase that was safe for use in food production (Ke et al., 2018), and *Citeromyces matritensis,* an osmotolerant, ethanol-producing yeast shown to be capable of ethanol production from salted algae (Okai et al., 2017).

Peru

Peru's culturable soil yeast population, consisting of 39 species, ranked the highest among sampled countries in Shannon diversity index. This population was unique in structure and composition as it contained 20 species not found in any other sampled country, including often misidentified pathogen and crude palm oil assimilator *Candida palmioleophila* (Jensen and Arendrup, 2011; NAKASE et al., 1988), rare pathogen *Filobasidium magnum* (Aboutalebian et al., 2020), and halotolerant yeast used in azo dye decolorization *Pichia occidentalis* (Wang et al., 2020). This population contained significantly more ascomycete species (n=29) than basidiomycete species (n=10). Peruvian population was notable for its relative evenness with no single species exceeding 12% in abundance. *Candida tropicalis* was the most prevalent (12%), followed by *Schwanniomyces occidentalis* (11%) and *Papiliotrema laurentii* (7%).

Saudi Arabia

Saudi Arabian culturable soil yeast population was the least diverse of all sampled countries according to the Shannon diversity index. Overall, the Saudi Arabian soil yeast population consisted of six basidiomycete species and three ascomycete species. One of the species was novel. This population was notable for the overwhelming prevalence of the human pathogenic yeast, *Cryptococcus deneoformans* (78%), the causative agent of fatal fungal meningoencephalitis. The genotypes of Saudi Arabian *C. deneoformans* strains have been reported in an earlier study (Samarasinghe et al., 2019). This study was the first to report the environmental presence of *C. deneoformans* in a desert climate.

Pathogenic yeast species

Based on recent information on yeast trophism (Kurtzman et al., 2011; Opulente et al., 2019), the following 12 species were the most common opportunistic yeast pathogens of humans worldwide: *Candida albicans, Candida dubliniensis, Candida glabrata, Candida guilliermondii* (syn: *Meyerozyma guilliermondii*), *Candida krusei* (*Pichia kudriavzevii*), *Candida lusitaniae* (syn: *Clavispora lusitaniae*), *Candida parapsilosis, Candida orthopsilosis, Candida metapsilosis, Candida tropicalis, Cryptococcus neoformans* and *Cryptococcus deneoformans*. We found 220 strains belonging to eight of these species, accounting for 15% of our global yeast population (Figure 3). *C. tropicalis* was both the most abundant and most widespread with 117 isolates originating from Cameroon, Canada, Costa Rica, France and Peru. The 76 *C. deneoformans* isolates were exclusively found in Saudi Arabian soils. Additionally, seven *C. krusei* isolates, six *C. albicans* isolates, and one *C. glabrata* isolate were found in Costa Rica, five *C. parapsilosis* isolates were found in Costa Rica, France and Saudi arabia, four *C. lusitaniae* isolates were found in Canada and France, and four *C. orthopsilosis* isolates were found in Cameroon and Costa Rica. Soil samples from China, Iceland and New Zealand did not yield any common yeast pathogens.

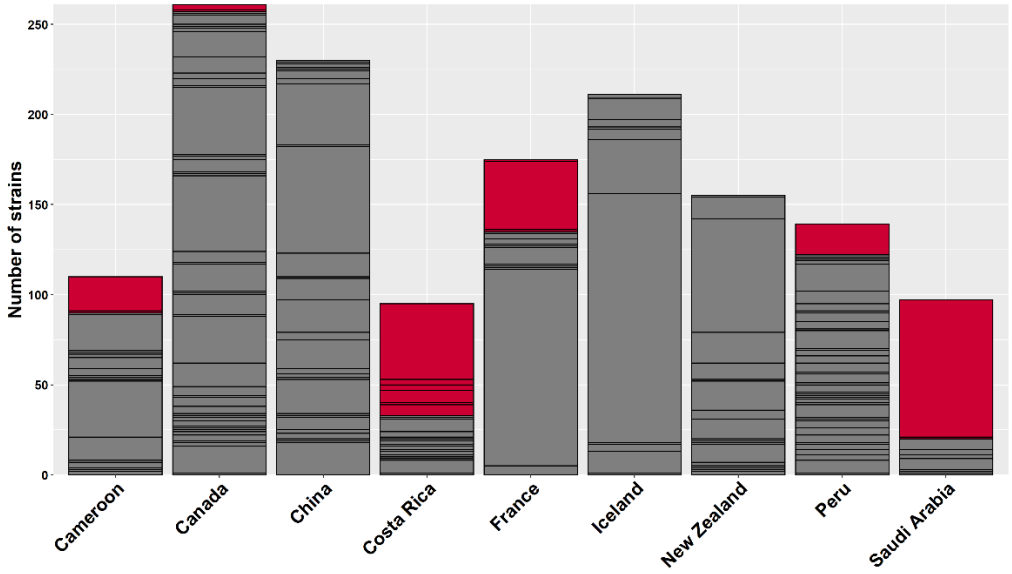


Figure 3: Pathogenic yeast species found in our global soil collection and their abundance by country. In these stacked bar plots, the pathogenic species are highlighted in red. The height of the red sections indicates their abundance. Soils from China, Iceland and New Zealand did not yield any pathogenic species.

Putative novel yeast species

Our yeast population included 44 potentially novel species from eight sampled countries: soil samples from Iceland did not yield any novel yeast species. Since we determined the most closely related genera for 41 species by running BLAST searches in the UNITE database (the remaining three species' ITS sequences were too short for analysis). Our 41 novel species can be categorized into 12 genera (9 in ascomycetes, 3 in basidiomycetes) with *Wickerhamomyces* containing ten potentially novel species, and *Candida* containing eight. For each genus, we constructed maximum likelihood (ML) trees using RaxML with 1000 bootstraps (Stamatakis, 2014) to determine the taxonomic placement of novel species with respect to all known species of that genus. Our ML trees confirmed the separation of newly discovered species from known species. For an example, the ML tree reveals the distinctness of the potentially novel *Wickherhamomyces* species we isolated in this study (Figure 4). We observed some geographical clustering where the Cameroonian and Canadian novel species formed their own clusters while the two novel species from Costa Rica and Peru clustered together. The ML trees of the remaining 11 genera can be found in Supplementary Dataset 1.

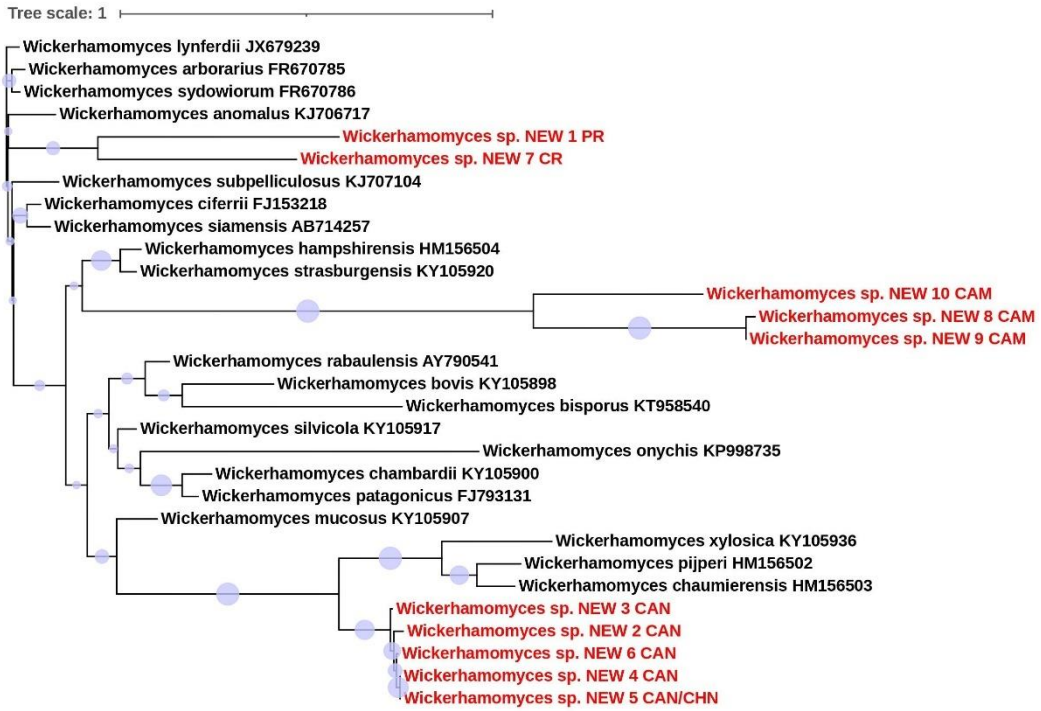


Figure 4: The maximum likelihood tree of *Wicherhamomyces* species based on rDNA ITS sequences. The placement of novel species with reference to known *Wickerhamomyces* species is shown. The novel species' country of origin is stated in the node labels where CAM = Cameroon, CAN = Canada and CHN = China. The tree was constructed using RaxML with 1000 bootstraps.

Predictors of global culturable soil yeast diversity

Our 47 distinct sampling locations covered a wide range of global climatic conditions (Table 2) with mean annual precipitation ranging from 0mm (Lima, Peru where there is virtually no rainfall) to 2965mm (Monteverde, Costa Rica), while mean annual temperature ranged from -1.4°C (Svartifoss, Iceland) to 29.6°C (Alqunfudah, Saudi Arabia). Elevation ranged from < 2m (some sites in Svartifoss, Iceland and Auckland, New Zealand) to 4922km above sea level (Rainbow Mountain, Peru). Mbandoumou in Cameroon was the closest to the equator (418.5km from equator) while Dimmuborgir in Iceland was the farthest (7280.5km from equator). Four locations, two in Saudi Arabia and two in Cameroon, were removed from further analysis as they did not yield any yeast isolates. The remaining 43 locations varied significantly in culturable yeast diversity as quantified by Shannon diversity index from 0 (only one species was found) to 2.77 (Fredericton, Canada). According to our mixed model, we found mean annual precipitation to be significantly correlated with the Shannon

diversity index ($p = 0.012$, Figure 5). We found no significant correlation between the remaining predictors and Shannon diversity index (Supplementary Dataset 3).

Table 2: Environmental and geographic characteristics of sampling sites. Mean annual rainfall, mean annual temperature, distance to equator and elevation of sampling sites are summarized here.

Country	City	Site code	Mean annual rainfall (mm)	Mean annual temperature (°C)	Distance to equator (km)	Altitude (meters above sea level)
Cameroon	Babanki	CBB	1813.312	20.75	20.75	1173
	Bambui	CBM	1813.312	20.75	20.75	1274
	Eloundem	CEL	1617.423	24.75	24.75	620
	Makepe	CMK	2825.192	26.83	26.83	62
	Mbalgon	CML	1617.423	25.03	25.03	556
	Mbandoumou	CMD	1617.423	24.22	24.22	719
	Mbingo	CMB	1813.312	20.75	20.75	1909
	Njinikejum	CNJ	1813.312	20.75	20.75	1573
	Simbock	CSM	1617.423	24.99	24.99	643
Canada	Fredericton	CF	1267.842	5.29	5.29	10
	Vancouver	CV	1567.708	9.95	9.95	31
China	Ailao Mountain	CAC	1120.492	17.29	17.29	2782
	Fenyi	CC	1556.404	17.26	17.26	307
	Jinfo Mountain	CJ	1154.85	15.49	15.49	2085
	Pangquangou Nature Reserve	CT	467.3346	9.66	9.66	775
	Taihang mountains west of Jincheng	CSX	515.0269	12.11	12.11	2012
Costa Rica	El Jardin	EJ	2308.358	27.23	27.23	152
	La Fotuna	LF	2964.685	25.51	25.51	392
	La Paz	LP	2964.685	25.51	25.51	1512
	Manuel Antonio	MA	2308.358	27.23	27.23	13
	Monteverde	MV	2964.685	25.51	25.51	1585
	Playa Hermosa	PH	2308.358	27.23	27.23	7
	Playa Samara Beach	SB	1682.769	26.72	26.72	7
	Poas Volcano	PV	2957.59	25.51	25.51	2350
	Samara Town	ST	1682.769	26.72	26.72	13
	Villas Playa Samara Beach Front Hotel	SH	1682.769	26.72	26.72	7
France	Hyeres	FHF	703.8039	15.45	15.45	37
	Uptown/Downtown Nice	FN	814.7077	13.82	13.82	25
Iceland	Dimmuborgir above Myvatn lake (highlands)	ID	969.2692	1.06	1.06	283
	Landbrotalaug mini hotspring	IL	823.4154	4.65	4.65	11
	National park near Svartifoss	ISF	1904.319	-1.4	-1.4	192
	Near Nautholsvik Geothermal Beach near Reykjavik university	IN	959.4077	4.17	4.17	0
	Skútustaðagígar pseudocraters on Myvatn lake	ISP	855.5423	1.93	1.93	271
	Thingvellir	IT	1107.946	3.05	3.05	89
New Zealand	Auckland	NZA	1172.973	15.13	15.13	0
Peru	Amazon	PA	2139.142	25.51	25.51	176
	Cusco	PC	679.0462	9.59	9.59	3322
	Lima	PL	0	19.57	19.57	138
	Machu Pichu	PM	563.1885	9.5	9.5	1940
	Rainbow Mountain	PR	725.7192	5.77	5.77	4922
	Sacred Valley	PS	636.3077	8.52	8.52	2866
Saudi Arabia	Alqunfudah	SAA	49.56154	29.58	29.58	1
	Dammam	SAD	73.69615	27.15	27.15	5
	Jeddah	SAJ	45.29615	29.25	29.25	15
	Medina	SAM	65.06154	27.18	27.18	636
	Umluj	SAU	4.815385	27.63	27.63	14
	Yanbu	SAY	29.56923	27.69	27.69	9

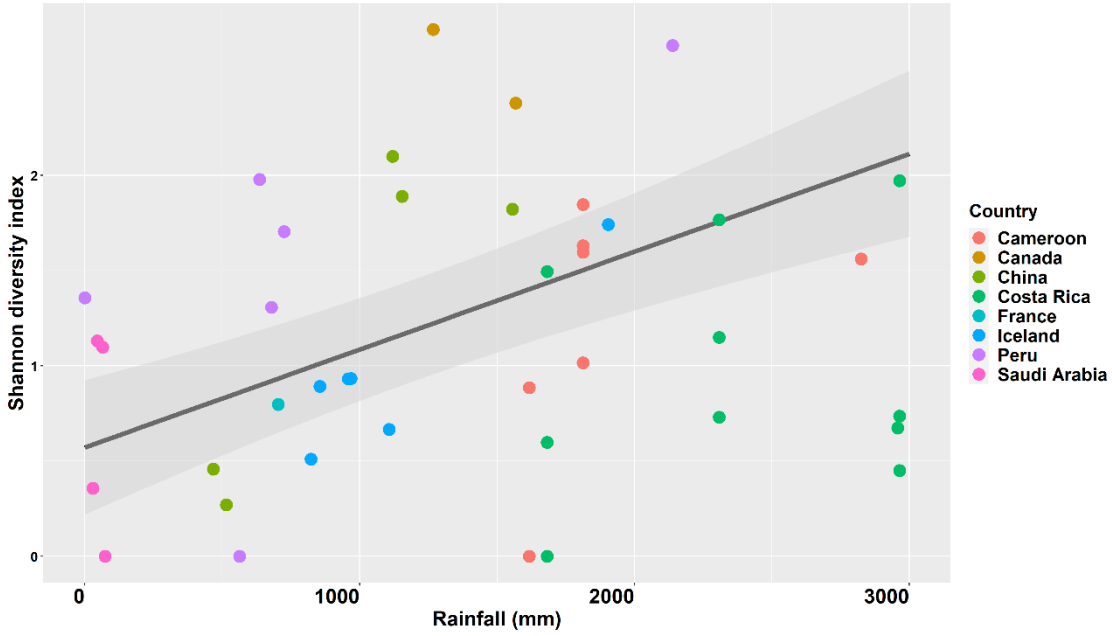


Figure 5: Mean annual precipitation is significantly correlated with soil yeast diversity. Here, the Shannon diversity index of our sampling sites is plotted against mean annual precipitation. Sampling sites are colored by country. The line plots the model predictions with associated uncertainty shaded in grey.

Air traffic volume as a predictor of shared species between countries

36 yeast species were found in more than one country. The following five country pairs had no soil yeast species in common: Iceland-Cameroon, Iceland-Costa Rica, Iceland-France, Iceland-Saudi Arabia, and Cameroon-Saudi Arabia. The number of shared species between the remaining 31 pairs ranged from one to 11 (Figure 6b). Air traffic volume data extracted from the Global Transnational Mobility Dataset showed that 25 700 496 trips were made between China and France between 2011-2016. During the same period, only 81 trips were made between Iceland and Cameroon (Figure 6a, Supplementary Table 3). We performed a linear regression analysis between air traffic volume, geographic distance, and the number of shared species between countries. While we did not find a significant correlation between geographic distance and the number of shared species, air traffic volume was significantly correlated with the number of shared species between countries ($p = 0.003$, Figure 7).

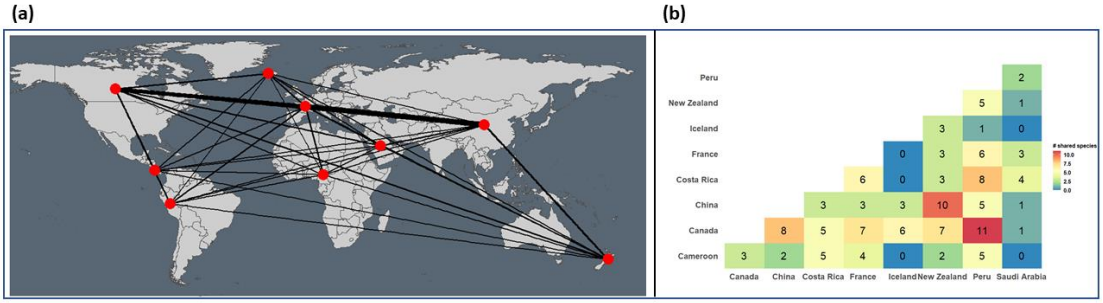


Figure 6: Air traffic volume between countries is correlated with number of shared species. (a) Volume of air traffic between the nine countries from 2011-2016. Thickness of the line indicates volume. (b) Heat map showing the number of shared species between country pairs.

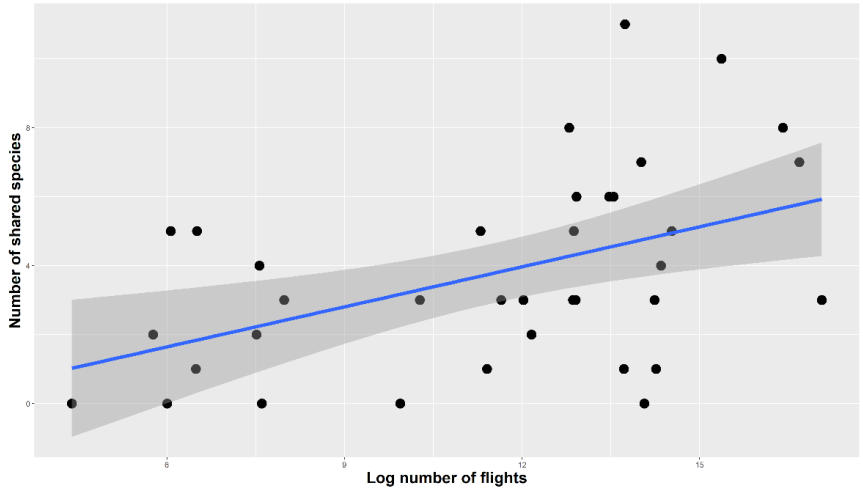


Figure 7*:* Number of shared species between countries is significantly correlated with traffic volume between them. We plotted the log number of trips made between country pairs between 2011-2016 against the number of yeast species shared between them.

We further constructed neighbour-joining (NJ) trees in MEGA7 (Kumar et al., 2016) based on ITS sequences of the four most shared species in our yeast population: *ebaryomyces hansenii* (7 countries), *Papiliotrema laurentii* (6 countries), *Candida tropicalis* (5 countries) and *Torulaspora delbrueckii* (5 countries). The NJ trees highlighted the lack of strict geographical clustering of isolates by country: for an example, most *P. laurentii* isolates found in New Zealand, Costa Rica, China, Peru, and Cameroon had identical ITS sequences and formed a cluster with most Canadian isolates (Figure 7). The NJ trees for the remaining three species can be found in Supplementary Dataset 2. This result is consistent with the hypothesis of recent long-distance dispersals for many of the shared species.

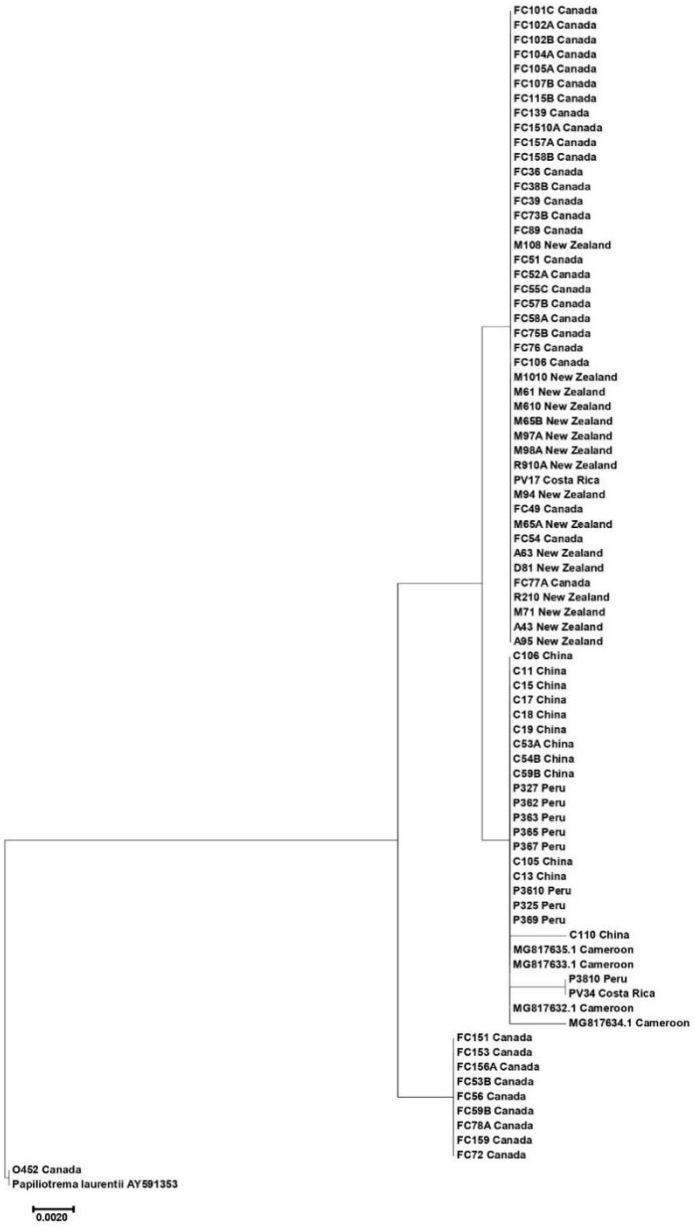


Figure 8: The neighbour-joining tree of *Papiliotrema laurentii* global isolates found in our study. Limited geographical clustering is observed, suggesting frequent gene flow between populations.

Comparison to culture-independent, metagenomics approach

We compared our results to a culture-independent study conducted by Tedersoo and colleagues who used high-throughput sequencing to investigate soil fungal diversity in 39 countries (Tedersoo et al., 2014). We detected 146 OTUs among our global yeast dataset based on the clustering of the ITS2 region at 98.41% sequence identity, while the metagenomics study reported a total of 50,589 fungal OTUs, with the number of yeast OTUs not known. While we were able to annotate all obtained ITS sequences to species level or a higher taxonomic status within the fungal kingdom, ~33% of the fungal OTUs in the metagenomics dataset were merely annotated as Environmental_sequence (724, 1.4%), uncultured_soil_fungus (2405, 4.8%), uncultured_ectomycorrhizal_fungus (1407, 2.8%) or uncultured_fungus (11898, 23.5%). BLAST searches revealed that 26% of our OTUs (38/146) had a significant match (>98.41% identity) in the metagenomics dataset: only 3% (4/38) matched a fungal OTU from the same country, while the remaining 21% matched a fungal OTU found in a different country.

2.6 Discussion

Despite being one of the most accessible ecological niches, soil remains an enigmatic source of yeast diversity and ecology. Given that most yeast species are not geographically widely distributed, extensive environmental sampling across diverse regions, habitats and climates is required to uncover new species and diversity patterns. Elucidating global patterns and trends would also allow us to predict the structure and diversity of soil yeast populations in unsampled locations. Using a set of global soil samples from nine countries in six continents, we address this knowledge gap by characterizing global patterns and potential predictors of culturable soil yeast diversity. Our study uncovered 134 soil yeast species among 1473 isolates, including 41 putative novel species. We identified mean annual precipitation and air traffic volume as significant predictors of soil yeast communities on a global scale. Our findings highlight the influence of both climatic factors and anthropogenic activity on soil yeast populations across the globe.

We found mean annual precipitation to be the strongest predictor of culturable soil yeast diversity across both local and global scales. Previous metagenomic studies have

established mean annual precipitation as one of the climatic drivers of soil fungal diversity (Egidi et al., 2019; Tedersoo et al., 2014). Our results confirm that this trend persists for culturable yeast communities in global soils as well. A previous study showed that vegetation was not a strong predictor of soil fungal diversity (Tedersoo et al., 2014). However, factors influencing soil yeast diversity may be different from the overall soil fungal diversity. For example, many yeast species isolated here have been reported as associated with plant materials. Both *P. kudriavzevii and M. guilliermondii* are common fruit and leaf inhabitants. *C. tropicalis* and *C. orthopsilosis* are frequently associated with rotting wood and plant materials in different ecosystems (Opulente et al., 2019). Thus, the composition and structure of soil yeast communities may reflect their associations with plants and plant-related substrates present in these soils. In addition, a few yeast species such as *C. tropicalis, C. saturnus, D. hansenii, P. laurentii, S. terrea,* and *S. terricola,* are cosmopolitan and known to be geographically broadly distributed (Kurtzman et al., 2011).

Earlier studies have shown that fungal communities in dry, semi-arid soils contain significantly more Ascomycota fungi than Basidiomycota (Abed et al., 2013; Murgia et al., 2019; Suleiman et al., 2019). We found a reversal of this trend in global soil yeast communities where basidiomycetous yeasts were found to be more prevalent in sampling sites receiving less rainfall ($p < 0.05$). Some soil-dwelling, basidiomycetous yeasts are known to produce biofilms that allow them to persist in low moisture, oligotrophic conditions (Spencer and Spencer, 1997). Low moisture, and resulting lack of nutrients, could favor cellular structures and metabolic activities of yeasts in one Phylum over the other, creating rainfall-associated global diversity patterns observed in our study. Given our findings, we hypothesize that extreme rainfall and drought events brought on by global warming can potentially shift the established landscape of soil yeast communities. This is especially alarming given the significant presence of pathogenic yeasts we detected in the soils. Specifically, 15% of all yeast isolates found in our study belong to common opportunistic pathogenic yeast species capable of causing deadly systemic infections. Altered rainfall patterns, and resulting changes in soil microclimates, could cause outgrowths of pathogenic species and lead to emergence of new fungal infections. With soil ecosystems being an important

source of bacterial and fungal infections, any changes and shifts in soil microbiomes could pose a significant threat to global public health.

Each of the nine countries investigated in our study was unique in the composition and structure of its culturable soil yeast population. 73% of the yeast species found in our study (98 out of 134) were specific to a single country. The fragmented nature of soil yeast distributions has been noted in previous studies where only a few species were found to be shared between sampling sites, even within the same region or country (Yurkov, 2018). The nine countries included in our study are separated by thousands of kilometers, with the two closest countries being France and Iceland (2235 km). Geographic isolation and local ecological conditions are likely key factors in limiting the spread and successful colonization of yeasts between populations, at least until recently when anthropogenic activity has strongly improved the connectivity between countries and continents and altered local ecological niches.

Our findings suggest that human activities have likely contributed to changing yeast distributions in soil environments across the globe. International travel has increased exponentially in the past few decades with international tourist arrivals increasing from 25 million in 1950 to a record-high 1.4 billion in 2018 (UNWTO, 2018). The global air transportation network has the small-world property where most countries can be reached from each other via a few flight hops (Wandelt and Sun, 2015). While we found unique yeast species in most localities, 36 yeast species (~25% of all species found) were shared between at least two countries. While the presence of some species in multiple countries could be due to natural dispersal events such as wind throughout that species' history, the correlation we found between air traffic volume and the number of shared species between countries suggested that human travel might have facilitated the spread of at least some yeasts across geographical borders. The lack of ITS sequence-based geographical clustering of the most shared species in our population is also consistent with gene flow between yeast populations in different countries. For *C. tropicalis*, the conclusion of gene flow among geographic populations is supported by a recent analysis of DNA sequences at six gene fragments for 876 global isolates that revealed sharing of multilocus genotypes between countries and continents (Wu et al., 2019). However, for other yeast species shared

between countries and continents, their detailed phylogeographic patterns remain to be elucidated. The COVID-19 pandemic has greatly shifted the political and economic landscape of our planet. Tourism both within and between countries has seen a drastic drop with accompanying tightening of borders between countries. The potential impact of the COVID-19 pandemic on culturable yeast populations remains to be determined.

Many yeast species have been recently reported and many more remain to be discovered (Kurtzman et al., 2011; Wu et al., 2019; Yurkov, 2018; Naranjo-Ortiz and Gabaldón, 2019). Both culture-dependent and independent studies routinely discover novel yeast species from the environment. Our study is one of many recent surveys to find previously undescribed species accounting for as much as 30% of natural yeast populations (Yurkov, 2018), implying that every one of three yeast species recovered from the environment is likely to be a new one. In fact, our results on soil yeast species diversity described here likely represent underestimates of both the total species in these soils as well as the percentage of novel yeast species in the soil. For example, our media and incubation conditions favor fast-growing yeasts under the selected conditions. If both fast-growing and slow-growing yeasts were in the same 0.1 gram of soil, it is likely that we would have recovered the fast-growing ones only due to its growth advantage.

The isolated pure yeast cultures could be of potential applied significance. For example, investigators often turn to natural soils in search of novel yeast strains with commercial and biotechnological potential. The metabolic and fermentative capabilities of the novel yeasts such as those in genera *Kazachstania* and *Blastobotrys*, found in Peruvian and French soils respectively, and yeasts like *Pichia kudriavzevii* (syn. *Candida krusei*) found in Costa Rican soil remain to be evaluated. In 2018, a novel strain of *Pichia kudriavzevii* isolated from soil in a sugarcane field in Thailand was shown to be more thermotolerant and produce more ethanol than the Thai industrial strain *Saccharomyces cerevisiae* TISTR 5606 (Pongcharoen et al., 2018). Presence of species of the genus *Kazachstania* in mixed cultures of *Saccharomyces cerevisiae* gives rise to fermented wines with diverse aroma profiles: however, *Kazachstania* species are unable to complete fermentation in monocultures (Jood et

al., 2017). Discovery of new *Kazachstania* species with more desirable fermentative abilities can aide the full exploitation of this genus in commercial wine fermentation. The thermo and halotolerant yeast *Blastobotrys adeninivorans* aids in a wide range of biotechnological applications including the production of secretory enzymes, as a host for heterologous gene expression and as a biological component in biosensors (Kunze et al., 2017).

In recent years, researchers have come to view metagenomics as a valuable tool in the investigation of microbial diversity in complex ecological systems. The large amount of data generated from high throughput sequencing is crucial for unearthing large-scale patterns at higher taxonomic levels. For example, a 2015 study compared culture-independent vs. culture-dependent characterizations of microbes from hydrocarbon-contaminated soil and found that the two methodologies provided diverging views of microbial communities, with only 8.2% of the fungal OTUs being shared between the two datasets (Stefani et al., 2015). A meta-analysis from 2019 estimated the total number of fungal species to be 7.8-8.8 times that of culturable species (Wu et al., 2019). However, limited information on yeast diversity could be extracted from previous metagenomics studies on global soil fungal diversity (Egidi et al., 2019; Tedersoo et al., 2014). Yeasts cannot be easily identified based on sequences alone: extracting ITS sequences of known yeasts from large metagenomics datasets is a time-consuming task that requires personnel with advanced knowledge of yeast taxonomy. For potentially novel species, the metagenomic approach would completely fail to identify them as yeasts, as they rely on the current state of knowledge on species taxonomy and annotation. In Tedersoo et al.'s study (2014), over 30% of the fungal OTUs remained minimally annotated. In the future, it may be possible to design multiple sets of primers to target known groups of yeasts through the metagenomic approach while excluding their respective closely related groups of filamentous fungi. However, even with this approach, only those closely related to known yeasts will be amplified while novel yeasts may still be missed. Here, using culture-dependent methods, we succeeded in isolating many soil yeasts, which aided in the characterization of global patterns of soil yeast distribution and diversity. Our global soil yeast collection with identity established and manually validated via ITS

sequencing provides a much-needed reference set for future investigators on yeast diversity and taxonomy.

Fungi isolated via culture-dependent methods can be identified as yeasts by morphology and can be further characterized using genomics, metabolomics, and transcriptomics (Xu 2020). Given the relatively low numbers of yeast cells in soil compared to bacteria, mold and other fungi, their DNA can easily escape detection in metagenomics studies, which could explain the lack of overlapping yeast sequences between our study and that of Tedersoo et al. (2014). On average, 40% of fungal DNA in soil are extracellular or comes from cells that were no longer intact, causing estimates of fungal richness to be Inflated by as much as 55% (Carini et al., 2016). Selective enrichment and culturing from soil samples in the lab remains the most effective ways of identifying and studying yeasts.

2.7 Conclusions

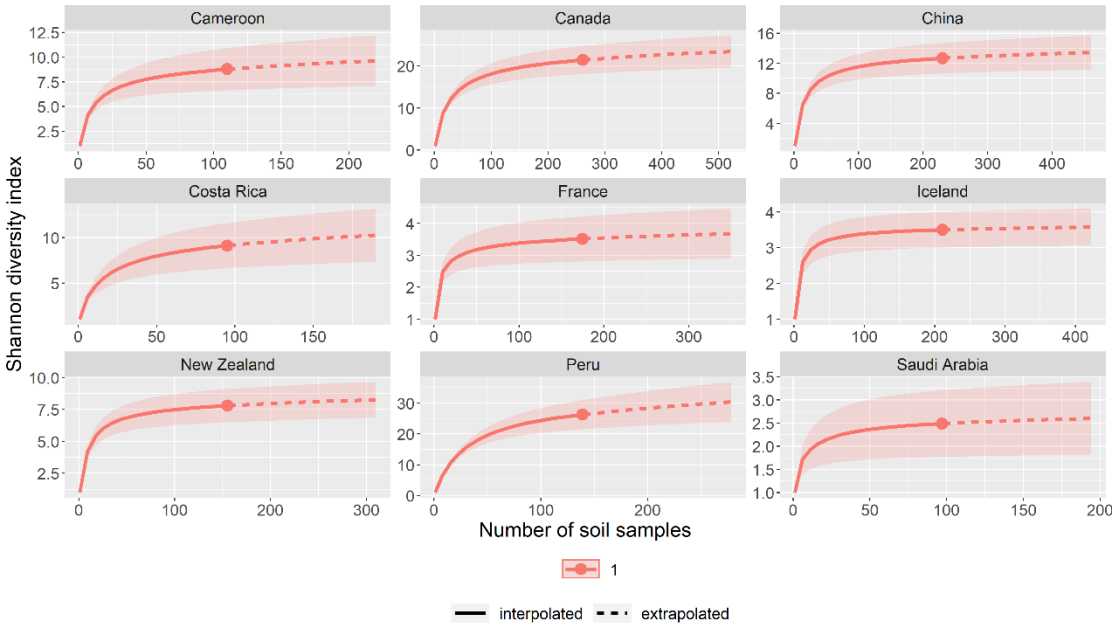
Our investigation into global patterns in culturable soil yeast diversity reaffirms soil as an important reservoir of environmental yeast species, both known and yet undiscovered. Precipitation emerges as the main predictor of soil yeast diversity across local and global scales. Ongoing global warming crisis and accompanying changes in rainfall could lead to expansion of pathogenic yeasts that already account for a sizable proportion of soil yeast communities. Our findings point to international travel being a potential contributing factor to the movement of yeast species across borders, with phylogenetic evidence suggesting long-distance gene flow between yeast populations. More environmental sampling is required to further uncover soil yeast diversity, isolates with commercial and biotechnological value and to monitor species that could pose a threat to human health.

2.8 Acknowledgements

We thank the following people for contributing soil samples: Haoran Jia, Yongjie Zhang, Thomas Harrison, Sebastian Harrison, Arshia Kazerouni, and Greg Korfanty. HS is supported by an NSERC CGS-D Scholarship. This research is supported by Natural Sciences and Engineering Research Council (NSERC) of Canada (RGPIN-

2020-05732) and by the Global Science Initiative Award of McMaster University to JX.

2.9 Supplementary Data



Supplementary Figure 1: Rarefaction curves showing saturation of Shannon diversity index

Supplementary Table 1

Year	Month	Country	City	Province	Latitude	Longitude	Site code
2016	June	Cameroon	Babanki	Northwest	6.12	10.26	CBB
2016	June	Cameroon	Bambui	Northwest	6.04	10.23	CBM
2016	June	Cameroon	Eloundem	Center	5.4	11.84	CEL
2016	June	Cameroon	Makepe	Littoral	4.14	9.83	CMK
2016	June	Cameroon	Mbalgon	Center	4.44	11.9	CML
2016	June	Cameroon	Mbandoumou	Center	3.77	11.51	CMD
2016	June	Cameroon	Mbingo	Northwest	6.16	10.64	CMB
2016	June	Cameroon	Njinikejum	Northwest	6.25	10.29	CNJ
2016	June	Cameroon	Simbock	Center	4.63	12.04	CSM
2019	August	Canada	Fredericton	New Brunswick	45.96	-66.64	CF
2020		Canada	Vancouver	British Columbia	49.28	-123.12	CV
2018		China	Ailao Mountain	Yunnan	24.21	101.32	CAC
2017		China	Fenyi	Jiangxi Province	27.48	114.41	CC
2019		China	Jinfo Mountain		29.01	107.17	CJ
		China	Pangquangou Nature Reserve, Jiaocheng County, Lüliang	Shanxi	37.55	112.56	CT
2017		China	Taihang mountains west of Jincheng		35.51	111.882	CSX
2019	March	Costa Rica	El Jardin	Alajuela	9.858	-84.576	EJ
2019	March	Costa Rica	La Fotuna	Alajuela	10.505	-84.687	LF
2019	March	Costa Rica	La Paz	Alajuela	10.204	-84.161	LP
2019	March	Costa Rica	Manuel Antonio	Puntarenas	9.39	-84.147	MA
2019	March	Costa Rica	Monteverde	Puntarenas	10.337	-84.81	MV
2019	March	Costa Rica	Playa Hermosa	Puntarenas	9.566	-84.59	PH
2019	March	Costa Rica	Playa Samara Beach	Guanacaste	9.873	-85.51	SB
2019	March	Costa Rica	Poas Volcano	Alajuela	10.197	-84.232	PV
2019	March	Costa Rica	Samara Town	Guanacaste	9.883	-85.528	ST
2019	March	Costa Rica	Villas Playa Samara Beach Front Hotel	Guanacaste	9.873	-85.509	SH
2019	August	France	Hyeres		43.12	6.12	FHF
2019	August	France	Nice		43.71	7.26	FN
2019	June	Iceland	Dimmuborgir above Myvatn lake (highlands)		65.59	-16.91	ID
2019	June	Iceland	Landbrotalaug mini hotspring		64.49	-22.19	IL
2019	June	Iceland	National park near Svartifoss		64.02	-16.97	ISF
2019	June	Iceland	Near Nautholsvik Geothermal Beach near Reykjavik university		64.12	-21.93	IN
2019	June	Iceland	Skútustaðagígar pseudocraters on Myvatn lake		65.56	-17.03	ISP
2019	June	Iceland	Thingvellir		64.25	-21.12	IT
2017		New Zealand	Auckland Domain	Auckland	-36.84	174.76	NZA
2017		New Zealand	Aut Millemium field	Auckland	-36.84	174.76	NZA
2017		New Zealand	Mount Eden	Auckland	-36.84	174.76	NZA
2017		New Zealand	Railway Station	Auckland	-36.84	174.76	NZA
2017		New Zealand	Trusts Arena Henderson	Auckland	-36.84	174.76	NZA
2017		New Zealand	Univ Auckland Campus	Auckland	-36.84	174.76	NZA
2019	May	Peru	Amazon		-12.52	-68.97	PA
2019	May	Peru	Cusco		-13.53	-71.96	PC
2019	May	Peru	Lima		-12.04	-77.04	PL
2019	May	Peru	Mauch Pichu		-13.15	-72.54	PM
2019	May	Peru	Rainbow Mountain		-13.86	-71.3	PR
2019	May	Peru	Sacred Valley		-13.33	-72.08	PS
2016	December	Saudi Arabia	Alqunfudah	Southwest	19.12	41.07	SAA
2016	December	Saudi Arabia	Dammam	East	26.39	49.97	SAD
2016	December	Saudi Arabia	Jeddah	West	21.28	39.23	SAJ
2016	December	Saudi Arabia	Medina	West	24.52	39.56	SAM
2016	December	Saudi Arabia	Umluj	Northwest	25.05	37.26	SAU
2016	December	Saudi Arabia	Yanbu	West	24.02	38.19	SAY

Supplementary Table 2: Soil yeast species discovered in this study, location of isolation and taxonomic annotations

Please click on the following link to access the table. https://docs.google.com/spreadsheets/d/1GRBWeywfFyWA97f75EOXdSGlywgED2vd/edit?usp=sharing&ouid=106686285052112798692&rtpof=true&sd=true

Supplementary Table 3: Air traffic volume between countries from 2011 to 2016

Country 1	Country 2	Total trips between 2011 and 2016
Cameroon	Canada	114066
Cameroon	China	190912
Cameroon	Costa Rica	430
Cameroon	France	1698224
Cameroon	Iceland	81
Cameroon	New Zealand	319
Cameroon	Peru	670
Cameroon	Saudi Arabia	20779
Canada	China	13341486
Canada	Costa Rica	2037266
Canada	France	17580898
Canada	Iceland	708188
Canada	New Zealand	1217830
Canada	Peru	923654
Canada	Saudi Arabia	908977
China	Costa Rica	166430
China	France	25700496
China	Iceland	401540
China	New Zealand	4734008
China	Peru	389346
China	Saudi Arabia	1561938
Costa Rica	France	408267
Costa Rica	Iceland	405
Costa Rica	New Zealand	28864
Costa Rica	Peru	360709
Costa Rica	Saudi Arabia	1924
France	Iceland	1281216
France	New Zealand	384373
France	Peru	763702
France	Saudi Arabia	1521833
Iceland	New Zealand	2920
Iceland	Peru	658
Iceland	Saudi Arabia	2002
New Zealand	Peru	80642
New Zealand	Saudi Arabia	89866
Peru	Saudi Arabia	1829

Supplementary Dataset 1: GenBank accession numbers of ITS sequences obtained in this study

Please click on the following link to access the table.

https://docs.google.com/spreadsheets/d/1UXt151cjnc58eR2HgmkMpfmlZcPERmCv/edit?usp=sharing&ouid=106686285052112798692&rtpof=true&sd=true

Supplementary Dataset 2: Maximum likelihood trees of novel species

Please click on the following link to access the table.

https://drive.google.com/file/d/188gb200vxHqfHwxvGT68UU4okKN4b0J3/view?usp=sharing

Supplementary Dataset 3: Neighbour-joining trees of species shared among countries

Please click on the following link to access the table.

https://drive.google.com/file/d/1kvp8enUNfNehYWPkfX77T5z2pcQYT_1F/view?usp=sharing

Supplementary Dataset 4: Summary of mixed model between species diversity, climatic factors, and geographical factors

Please click on the following link to access the table.

https://drive.google.com/file/d/1itnubg07kyQCKm4CjD0QlraP2AjyYgU7/view?usp=sharing

2.10 References

Abbasian, F., Palanisami, T., Megharaj, M., Naidu, R., Lockington, R., Ramadass, K., 2016. Microbial diversity and hydrocarbon degrading gene capacity of a crude oil field soil as determined by metagenomics analysis. Biotechnol. Prog. 32, 638–648. https://doi.org/10.1002/btpr.2249

Abed, R.M.M., Al-Sadi, A.M., Al-Shehi, M., Al-Hinai, S., Robinson, M.D., 2013. Diversity of free-living and lichenized fungal communities in biological soil crusts of the Sultanate of Oman and their role in improving soil properties. Soil Biol. Biochem. 57, 695–705. https://doi.org/10.1016/j.soilbio.2012.07.023

Abia, A.L.K., Alisoltani, A., Keshri, J., Ubomba-Jaswa, E., 2018. Metagenomic analysis of the bacterial communities and their functional profiles in water and sediments of the Apies River, South Africa, as a function of land use. Sci. Total Environ. 616–617, 326–334. https://doi.org/10.1016/j.scitotenv.2017.10.322

Aboutalebian, S., Mahmoudi, S., Okhovat, A., Khodavaisy, S., Mirhendi, H., 2020. Otomycosis Due to the Rare Fungi Talaromyces purpurogenus, Naganishia albida and Filobasidium magnum. Mycopathologia 185, 569–575. https://doi.org/10.1007/s11046-020-00439-8

Adam, H., Groenewald, M., Mohan, S., Richardson, S., Bunn, U., Gibas, C.F.C., Poutanen, S., Sigler, L., 2009. Identification of a new species, *Candida subhashii* , as a cause of peritonitis. Med. Mycol. 47, 305–311. https://doi.org/10.1080/13693780802380545

Bates, D., Mächler, M., Bolker, B.M., Walker, S.C., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01

Białkowska, A.M., Szulczewska, K.M., Krysiak, J., Florczak, T., Gromek, E., Kassassir, H., Kur, J., Turkiewicz, M., 2017. Genetic and biochemical characterization of yeasts isolated from Antarctic soil samples. Polar Biol. 40, 1787–1803. https://doi.org/10.1007/s00300-017-2102-7

Botha, A., 2011. The importance and ecology of yeasts in soil. Soil Biol. Biochem. https://doi.org/10.1016/j.soilbio.2010.10.001

Connell, L., Redman, R., Craig, S., Scorzetti, G., Iszard, M., Rodriguez, R., 2008. Diversity of soil yeasts isolated from South Victoria Land, Antarctica. Microb. Ecol. 56, 448–459. https://doi.org/10.1007/s00248-008-9363-1

Egidi, E., Delgado-Baquerizo, M., Plett, J.M., Wang, J., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K., 2019. A few Ascomycota taxa dominate soil fungal communities worldwide. Nat. Commun. 10, 1–9. https://doi.org/10.1038/s41467-019-10373-z

Findlater, A., Bogoch, I.I., 2018. Human Mobility and the Global Spread of Infectious Diseases: A Focus on Air Travel. Trends Parasitol. https://doi.org/10.1016/j.pt.2018.07.004

Frac, M., Hannula, S.E., Belka, M., Jędryczka, M., 2018. Fungal biodiversity and their role in soil health. Front. Microbiol. https://doi.org/10.3389/fmicb.2018.00707

Groenewald, M., Lombard, L., de Vries, M., Lopez, A.G., Smith, M., Crous, P.W., 2018. Diversity of yeast species from Dutch garden soil and the description of six novel Ascomycetes. FEMS Yeast Res. 18, 76. https://doi.org/10.1093/femsyr/foy076

Hilber-Bodmer, M., Schmid, M., Ahrens, C.H., Freimoser, F.M., 2017. Competition assays and physiological experiments of soil and phyllosphere yeasts identify Candida subhashii as a novel antagonist of filamentous fungi. BMC Microbiol. 17, 1–15. https://doi.org/10.1186/s12866-016-0908-z

Hsieh, T.C., Ma, K.H., Chao, A., 2016. iNEXT: an R package for rarefaction and extrapolation of species diversity (<scp>H</scp> ill numbers). Methods Ecol. Evol. 7, 1451–1456. https://doi.org/10.1111/2041-210X.12613

Into, P., Pontes, A., Sampaio, J.P., Limtong, S., 2020. Yeast Diversity Associated with the Phylloplane of Corn Plants Cultivated in Thailand. Microorganisms 8, 80. https://doi.org/10.3390/microorganisms8010080

Jensen, R.H., Arendrup, M.C., 2011. Candida palmioleophila: Characterization of a previously overlooked pathogen and its unique susceptibility profile in comparison with five related species. J. Clin. Microbiol. 49, 549–556.

https://doi.org/10.1128/JCM.02071-10

Jood, I., Hoff, J.W., Setati, M.E., 2017. Evaluating fermentation characteristics of Kazachstania spp. and their potential influence on wine quality. World J. Microbiol. Biotechnol. 33. https://doi.org/10.1007/s11274-017-2299-1

Kahle, D., Wickham, H., 2013. ggmap: Spatial Visualization with ggplot2. R J. 5, 144–161.

Ke, Q., Fulmer, P., Mizutani, A., 2018. Toxicological evaluation of β-Galactosidase enzyme produced by Papiliotrema terrestris. Regul. Toxicol. Pharmacol. 92, 213–219. https://doi.org/10.1016/j.yrtph.2017.12.002

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33, 1870–1874. https://doi.org/10.1093/molbev/msw054

Kunze, G., Bischoff, F., Chamas, A., Litwinska, K., Matthes, F., Boer, E., 2017. Applications of blastobotrys (arxula) adeninivorans in biotechnology, in: Yeast Diversity in Human Welfare. Springer Singapore, pp. 455–479. https://doi.org/10.1007/978-981-10-2621-8_18

Kurtzman, C., Fell, J.W., Boekhout, T. (Eds.), 2011. The Yeasts A Taxonomic Study, 5th edition. ed. Elsevier.

Li, A.H., Yuan, F.X., Groenewald, M., Bensch, K., Yurkov, A.M., Li, K., Han, P.J., Guo, L.D., Aime, M.C., Sampaio, J.P., Jindamorakot, S., Turchetti, B., Inacio, J., Fungsin, B., Wang, Q.M., Bai, F.Y., 2020. Diversity and phylogeny of basidiomycetous yeasts from plant leaves and soil: Proposal of two new orders, three new families, eight new genera and one hundred and seven new species. Stud. Mycol. 96, 17–140. https://doi.org/10.1016/j.simyco.2020.01.002

Li, X., Qin, L., 2005. Metagenomics-based drug discovery and marine microbial diversity. Trends Biotechnol. 23, 539–543. https://doi.org/10.1016/j.tibtech.2005.08.006

Monteiro Moreira, G.A., Martins do Vale, H.M., 2020. Soil Yeast Communities in

Revegetated Post-Mining and Adjacent Native Areas in Central Brazil. Microorganisms 8, 1116. https://doi.org/10.3390/microorganisms8081116

Murgia, M., Fiamma, M., Barac, A., Deligios, M., Mazzarello, V., Paglietti, B., Cappuccinelli, P., Al-Qahtani, A., Squartini, A., Rubino, S., Al-Ahdal, M.N., 2019. Biodiversity of fungi in hot desert sands. Microbiologyopen 8, e00595. https://doi.org/10.1002/mbo3.595

NAKASE, T., ITOH, M., SUZUKI, M., KOMAGATA, K., KODAMA, T., 1988. Candida palmioleophila sp. nov., a yeast capable of assimilating crude palm oil, formerly identified as Torulopsis candida. J. Gen. Appl. Microbiol. 34, 493–498. https://doi.org/10.2323/jgam.34.493

Naranjo-Ortiz, M.A., Gabaldón, T., 2019. Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. Biol Rev Camb Philos Soc. 94, 2101-2137. doi: 10.1111/brv.12550.

Ogaki, M.B., Teixeira, D.R., Vieira, R., Lírio, J.M., Felizardo, J.P.S., Abuchacra, R.C., Cardoso, R.P., Zani, C.L., Alves, T.M.A., Junior, P.A.S., Murta, S.M.F., Barbosa, E.C., Oliveira, J.G., Ceravolo, I.P., Pereira, P.O., Rosa, C.A., Rosa, L.H., 2020. Diversity and bioprospecting of cultivable fungal assemblages in sediments of lakes in the Antarctic Peninsula. Fungal Biol. 124, 601–611. https://doi.org/10.1016/j.funbio.2020.02.015

Okai, M., Betsuno, A., Shirao, A., Obara, N., Suzuki, K., Takei, T., Takashio, M., Ishida, M., Urano, N., 2017. Citeromyces matritensis M37 is a salt-tolerant yeast that produces ethanol from salted algae. Can. J. Microbiol. 63, 20–26. https://doi.org/10.1139/cjm-2016-0259

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., Maintainer, H.W., 2020. Package "vegan" Title Community Ecology Package Version 2.5-7.

Opulente, D.A., Langdon, Q.K., Buh, K.V., Haase, M.A.B., Sylvester, K., Moriarty, R.V., Jarzyna, M., Considine, S.L., Schneider, R.M., Hittinger, C.T., 2019.

Pathogenic budding yeasts isolated outside of clinical settings. FEMS Yeast Res. 19, foz032. doi: 10.1093/femsyr/foz032.

Pongcharoen, P., Chawneua, J., Tawong, W., 2018. High temperature alcoholic fermentation by new thermotolerant yeast strains Pichia kudriavzevii isolated from sugarcane field soil. Agric. Nat. Resour. 52, 511–518. https://doi.org/10.1016/j.anres.2018.11.017

Recchi, E., Deutschmann, E., Vespe, M., 2019. Estimating Transnational Human Mobility on a Global Scale. SSRN Electron. J. https://doi.org/10.2139/ssrn.3384000

Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G., Triplett, E.W., 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J. 1, 283–290. https://doi.org/10.1038/ismej.2007.53

Samarasinghe, H., Aljohani, R., Jimenez, C., Xu, J., 2019. Fantastic yeasts and where to find them: The discovery of a predominantly clonal Cryptococcus deneoformans population in Saudi Arabian soils. FEMS Microbiol. Ecol. 95. https://doi.org/10.1093/femsec/fiz122

Spencer, J.F.T., Spencer, D.M., 1997. Ecology: Where Yeasts Live, in: Yeasts in Natural and Artificial Habitats. Springer Berlin Heidelberg, pp. 33–58. https://doi.org/10.1007/978-3-662-03370-8_4

Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stefani, F.O.P., Bell, T.H., Marchand, C., Providencia, I.E. de la, Yassimi, A. El, St-Arnaud, M., Hijri, M., 2015. Culture-Dependent and -Independent Methods Capture Different Microbial Community Fractions in Hydrocarbon-Contaminated Soils. PLoS One 10, e0128272. https://doi.org/10.1371/JOURNAL.PONE.0128272

Suleiman, M.K., Dixon, K., Commander, L., Nevill, P., Quoreshi, A.M., Bhat, N.R.,

Manuvel, A.J., Sivadasan, M.T., 2019. Assessment of the diversity of fungal community composition associated with Vachellia pachyceras and its Rhizosphere soil from Kuwait desert. Front. Microbiol. 10, 63. https://doi.org/10.3389/fmicb.2019.00063

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Ruiz, L.V., Vasco-Palacios, A.M., Thu, P.Q., Suija, A., Smith, M.E., Sharp, C., Saluveer, E., Saitta, A., Rosas, M., Riit, T., Ratkowsky, D., Pritsch, K., Põldmaa, K., Piepenbring, M., Phosri, C., Peterson, M., Parts, K., Pärtel, K., Otsing, E., Nouhra, E., Njouonkou, A.L., Nilsson, R.H., Morgado, L.N., Mayor, J., May, T.W., Majuakim, L., Lodge, D.J., Lee, S.S., Larsson, K.-H., Kohout, P., Hosaka, K., Hiiesalu, I., Henkel, T.W., Harend, H., Guo, L., Greslebin, A., Grelet, G., Geml, J., Gates, G., Dunstan, W., Dunk, C., Drenkhan, R., Dearnaley, J., De Kesel, A., Dang, T., Chen, X., Buegger, F., Brearley, F.Q., Bonito, G., Anslan, S., Abell, S., Abarenkov, K., 2014. Global diversity and geography of soil fungi. Science (80-.). 346.

Tepeeva, A.N., Glushakova, A.M., Kachalkin, A. V., 2018. Yeast communities of the Moscow city soils. Microbiol. (Russian Fed. 87, 407–415. https://doi.org/10.1134/S0026261718030128

UNWTO World Tourism Barometer and Statistical Annex, November 2018, 2018. . UNWTO World Tour. Barom. 16, 1–36. https://doi.org/10.18111/wtobarometereng.2018.16.1.5

Villarreal, P., Carrasco, M., Barahona, S., Alcaíno, J., Cifuentes, V., Baeza, M., 2018. Antarctic yeasts: Analysis of their freeze-thaw tolerance and production of antifreeze proteins, fatty acids and ergosterol. BMC Microbiol. 18, 66. https://doi.org/10.1186/s12866-018-1214-8

Vishniac, H.S., 2006. A multivariate analysis of soil yeasts isolated from a latitudinal gradient. Microb. Ecol. https://doi.org/10.1007/s00248-006-9066-4

Vishniac, H.S., 1996. Biodiversity of yeasts and filamentous microfungi in terrestrial Antarctic ecosystems. Biodivers. Conserv. 5, 1365–1378. https://doi.org/10.1007/BF00051983

Vu, D., Groenewald, M., Szöke, S., Cardinali, G., Eberhardt, U., Stielow, B., de Vries, M., Verkleij, G.J.M., Crous, P.W., Boekhout, T., Robert, V., 2016. DNA barcoding analysis of more than 9 000 yeast isolates contributes to quantitative thresholds for yeast species and genera delimitation. Stud. Mycol. 85, 91–105. https://doi.org/10.1016/j.simyco.2016.11.007

Wandelt, S., Sun, X., 2015. Evolution of the international air transportation country network from 2002 to 2013. Transp. Res. Part E Logist. Transp. Rev. 82, 55–78. https://doi.org/10.1016/j.tre.2015.08.002

Wang, X., Wang, Y., Ning, S., Shi, S., Tan, L., 2020. Improving Azo Dye Decolorization Performance and Halotolerance of Pichia occidentalis A2 by Static Magnetic Field and Possible Mechanisms Through Comparative Transcriptome Analysis. Front. Microbiol. 11, 712. https://doi.org/10.3389/fmicb.2020.00712

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. J. Open Source Softw. 4, 1686. https://doi.org/10.21105/joss.01686

Wu, B., Hussain, M., Zhang, W., Stadler, M., Liu, X., Xiang, M., 2019. Current insights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi. Mycology 10, 127–140. https://doi.org/10.1080/21501203.2019.1614106

Wu, J.Y., Zhou, D.Y., Zhang, Y., Mi, F., Xu, J. 2019. Analyses of the global multilocus genotypes of the human pathogenic yeast Candida tropicalis. Frontiers in Microbiology. 10, 900. DOI: 10.3389/fmicb.2019.00900

Xu, J., 2016. Fungal DNA barcoding. Genome. 59, 913-932.

https://doi.org/10.1139/gen-2016-0046

Yurkov, A.M., 2018. Yeasts of the soil - obscure but precious. Yeast 35, 369–378. https://doi.org/10.1002/yea.3310

Yurkov, A.M., Kemler, M., Begerow, D., 2012. Assessment of yeast diversity in soils under different management regimes. Fungal Ecol. 5, 24–35. https://doi.org/10.1016/j.funeco.2011.07.004

Yurkov, A.M., Röhl, O., Pontes, A., Carvalho, C., Maldonado, C., Sampaio, J.P., 2016a. Local climatic conditions constrain soil yeast diversity patterns in Mediterranean forests, woodlands and scrub biome. FEMS Yeast Res. 16, fov103. https://doi.org/10.1093/femsyr/fov103

Yurkov, A.M., Wehde, T., Federici, J., Schäfer, A.M., Ebinghaus, M., Lotze-Engelhard, S., Mittelbach, M., Prior, R., Richter, C., Röhl, O., Begerow, D., 2016b. Yeast diversity and species recovery rates from beech forest soils. Mycol. Prog. 15, 845–859. https://doi.org/10.1007/s11557-016-1206-8

Chapter 3

Fantastic Yeasts and Where to Find Them: The Discovery of a Predominantly Clonal *Cryptococcus deneoformans* Population in Saudi Arabian Soils

3.1 Preface

The geographical distribution of *C. deneoformans*, typically found in association with pigeon excreta, tree barks and soil, was previously believed to be largely limited to temperate climates. Its sister species *C. neoformans* was believed to be more thermotolerant with a worldwide distribution, including tropical regions. In the course of culturing yeasts from global soils, we isolated a large number of *C. deneoformans* strains from Saudi Arabian soil where it was the dominant yeast species in the population. Our study was the first to discover evidence of *C. deneoformans* persisting in desert climates. In this study, we reported this novel finding and genetically characterized the population by performing multilocus sequence typing on the strains. Our findings suggest that the geographical distribution of *C. deneoformans* needs to be expanded to include desert climates and that further sampling could reveal additional niches of this species.

This study is now published in FEMS Microbiology Ecology Volume 95, Issue 9. I am the primary contributor of this work. I conducted the majority of the experiments as well as the analyses and drafting of the manuscript. Co-authors of this article are Renad Aljohani, Carlene Jimenez and Jianping Xu.

3.2 Abstract

Cryptococcus deneoformans is an opportunist yeast pathogen and causative agent of meningoencephalitis in humans. It is known to be mainly distributed in temperate climates. Most of our current understanding of this species have come from clinical isolates, leaving environmental populations largely unexplored. The Middle East remains one such underexplored area with no published study to date investigating cryptococcal diversity in soil. In this study, we identified 76 *C. deneoformans* isolates from a survey of 562 soil samples collected from six cities in Saudi Arabia. Multilocus sequence typing revealed the presence of two major sequence types (ST), ST160 (n=63) and ST294 (n=9) along with four singleton STs, three of which were novel. One novel ST, ST613, was likely a recombinant product between ST160 and ST294. Among the 76 isolates, 75 belonged to mating type (MAT) α while one isolate was MATa. Our analyses suggest that the Saudi Arabian *C. deneoformans* population likely reproduces both asexually and sexually in nature. Our study is the first to report the occurrence of *C. deneoformans* in a desert climate, representing a novel expansion to this species' currently known ecological niche.

3.3 Introduction

The two sister species, *Cryptococcus neoformans* and *Cryptococcus deneoformans* are opportunistic human yeast pathogens and causative agents of cryptococcosis, a set of life-threatening diseases that includes respiratory infections, skin lesions and meningoencephalitis (Charlier et al. 2009; Goldman, Lee, and Casadevall 1994; Husain, Wagener, and Singh 2001; Del Valle and Pina-Oviedo 2006). Each year, over 220,000 cryptococcal infections are reported worldwide with 180,000 resulting in mortality (Rajasingham et al. 2017). Based on cell surface antigen properties and genetic differences, cryptococcal strains are categorized into serotypes and molecular types (Bhattacharjee, Bennett and Glaudemans 1984; Boekhout *et al.* 2001; Meyer *et al.* 2009). *C. neoformans* corresponds to serotype A and molecular types VNI, VNII, and VNB, while *C. deneoformans* belongs to serotype D and molecular type VNIV. The two lineages have been alternatively classified as two varieties of one species (namely *C. neoformans* var. *grubii* for serotype A, molecular types VNI, VNII, and VNB; and *C. neoformans* var. *neoformans* for serotype D, molecular type VNIV)

(Hagen et al. 2015, 2017; Kwon-Chung et al. 2017). In this paper, we will use the two-species nomenclature.

While sharing 85-90% sequence identity at the genomic level, the two species show characteristic differences in virulence factors such as melanin production, growth at 37°C and capsule production, where *C. neoformans* generally surpasses *C. deneoformans* in the expression of these phenotypes (Kavanaugh, Fraser, and Dietrich 2006; Martinez, Garcia-Rivera, and Casadevall 2001; Vogan et al. 2016). *C. neoformans* is responsible for the majority of global cryptococcal infections although cutaneous infections are predominantly associated with *C. deneoformans* strains (Barchiesi et al. 2005; Dromer et al. 1996; Pini, Faggi, and Bravetti 2017; Sanchini et al. 2014). Due to its higher thermotolerance, *C. neoformans* is frequently isolated from tropical and subtropical regions with strains having been discovered from environmental sources in all continents except Antarctica (Massimo Cogliati 2013). *C. deneoformans* appears to be less thermotolerant and its geographical distribution is largely contained to temperate climates (Massimo Cogliati 2013). Prevalence of *C. deneoformans* is highest in Mediterranean Europe with environmental and clinical isolates having been recovered from many countries including Croatia, Denmark, France, Germany, Greece, Italy, Russia, Serbia, Slovenia, Spain and Sweden (Arsenijevic et al. 2014; Massimo Cogliati et al. 2018; Massimo Cogliati, D'Amicis, et al. 2016; Desnos-Ollivier et al. 2015; Gago et al. 2017; Hagen et al. 2016; Mlinaric-Missoni et al. 2011; Pini, Faggi, and Bravetti 2017; Sanchini et al. 2014; Tomazin et al. 2017; Viviani et al. 2006) . While *C. deneoformans* causes <5% of cryptococcal infections on a global scale, it is responsible for over 30% of all cryptococcal infections in Europe (Arsenijevic et al. 2014; Viviani et al. 2006).

The current literature on *Cryptococcus* population structure and epidemiology is heavily biased towards clinical isolates. While surveys of clinical cryptococcal isolates, including their genotypes, phenotypes and clinical outcomes, are frequently reported from around the world, there are significantly fewer studies investigating the environmental presence of *Cryptococcus*. A meta-analysis conducted by Cogliati in 2013 that combined results from hundreds of global studies revealed that out of the 76,981 cryptococcal isolates reported in the literature, the overwhelming majority

were of clinical origin (n=70,198, 91.2%) while only 6,783 strains (8.8%) came from environmental sources (Massimo Cogliati 2013). This disparity is likely due to clinical samples being more easily accessible for investigation as well as the need for surveillance and monitoring for possible epidemics. However, since cryptococcal infections are non-contagious, hosts invariably acquire cryptococcal pathogens from the environment. Therefore, studying environmental populations will provide a more complete understanding of the true genetic diversity and virulence potential of the human pathogenic *Cryptococcus* in a region.

Due to its high prevalence, most studies on *Cryptococcus* have focused on *C. neoformans* with environmental surveys of *C. deneoformans* especially lacking. In fact, *C. deneoformans* isolates have never been reported from the African continent. Its apparent absence in the literature is likely due to insufficient sampling since interspecific hybrids between *C. neoformans* and *C. deneoformans* have been recovered from clinical and environmental sources in Africa (Kassi et al. 2016; A. P. Litvintseva et al. 2005; Miglia et al. 2011; Van Wyk et al. 2014). Another region that is scarcely surveyed for *Cryptococcus* is the Middle East, partly due to its desert climate being considered as not conducive for cryptococcal growth and survival. However, albeit rare, cryptococcal infections have been reported from several countries in the Middle East including Dubai, Kuwait, Qatar and Saudi Arabia (Khayhan et al. 2013; Kinne et al. 2017; Mansoor et al. 2015). Using multilocus sequence typing (MLST), a recent study concluded that the global *C. deneoformans* population has a recombing population structure with no clear geographical structuring (Massimo Cogliati, Zani, et al. 2016). This suggests that increases in human migration over the last several centuries may have facilitated the migration of *C. deneoformans* across the globe, blurring genetic boundaries among geographical populations and resulting in isolates being introduced to previously uninhabited niches. Therefore, it is important to survey previously unexplored environments for presence of *Cryptococcus*, with a focus on countries where cryptococcal infections have already been reported.

Saudi Arabia occupies almost 80% of the Arabian Peninsula and most of its territory consists of arid desert with a series of mountains extending through the Southwestern

and Western regions (Hegazy, El-Demerdash, and Hosni 1998; Vincent 2008). Except the Southwestern region, most of Saudi Arabia has a high annual temperature and low precipitation. A desert climate with high-temperature conditions increases evaporation and reduces moisture in soil, resulting in salt precipitation. Indeed, ground water in some places has a salt content comparable to that of sea water (Sallam 2002). At present, there is no published information on yeast diversity in Saudi Arabian soil. Furthermore, given these extreme conditions, Saudi Arabian soil is significantly different from currently known environmental habitats of cryptococcal species, although several human cryptococcal infections have been reported in Saudi Arabia over the years dating as far back as 1989 (Al-Hedaithy 1992).

In this study, we identified 114 yeast isolates from a survey of 562 soil samples collected from six cities in Saudi Arabia: 76 isolates were determined to be strains of *C. deneoformans*. We explored the genetic variation among these isolates using DNA sequence information obtained from the consensus MLST scheme for *Cryptococcus* species complex. Our analyses revealed for the first time a desert climate as an environmental reservoir of *C. deneoformans* and that this population has a predominantly clonal population structure with some evidence of sexual recombination.

3.4 Materials and Methods

Soil collection and yeast isolation: A total of 562 soil samples were collected from six Saudi Arabian cities in four regions: Yanbu, Jeddah and Al-medina on the West Coast by the Red Sea, Umluj in the Northwest, Al Qunfudah in the Southwest, and Dammam on the Persian Gulf Coast (Figure 1). The geographic coordinates of sampling locations are listed in Table 1. Samples were collected in December (coolest, dry season) of 2016. Each sample consisted of ~1 gram of soil collected at a depth of approximately 5 – 10 cm below the surface. Soil samples were at least five meters from each other in all directions. For yeast isolation, 0.1 g of soil was mixed into 1 ml of Sabouraud Dextrose Broth (SDB) containing 0.035 mg/ml of the antibiotic chloramphenicol and incubated at 30°C for 2 days. Samples were then sub-cultured onto Yeast Extract-Peptone-Dextrose agar (YEPD) and grown for 2 days at 30°C.

Colonies that appeared to be yeasts based on appearance and morphology (creamy or white color, smooth, matte texture) were selected for further investigation.

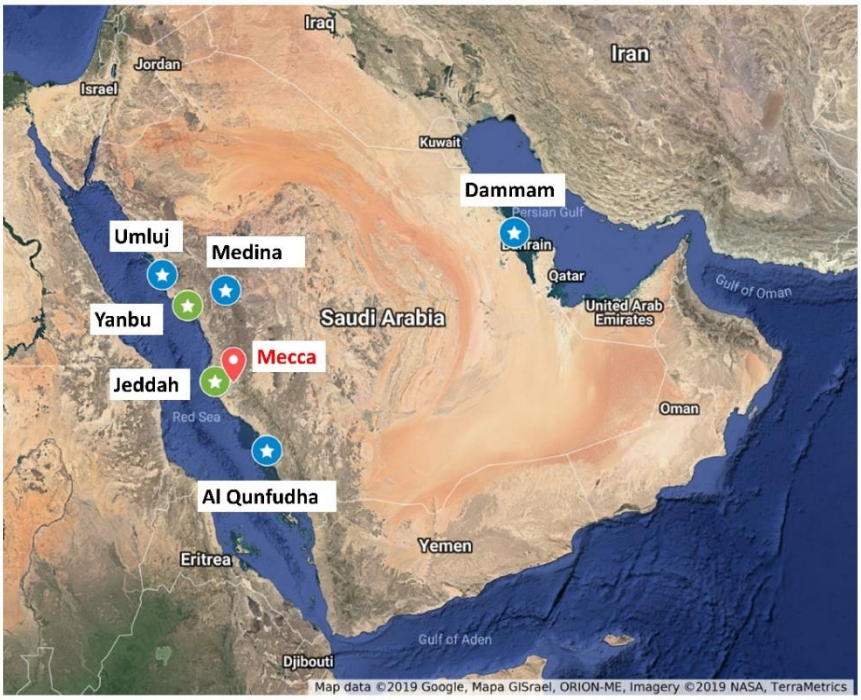


Figure 1: The sampling locations in Saudi Arabia. Soil samples were collected from six locations: Umluj, Medina, Yanbu, Jeddah, Al Qunfudha and Dammam. The two cities from which *C. deneoformans* isolates were recovered are highlighted in green. The city of Mecca is shown in red.

Table 1: The sampling locations in Saudi Arabia. A total of 562 soil samples were collected from six cities in Saudi Arabia. 114 yeast isolates were recovered 76 of which were determined to be *C. deneoformans*.

Region	City	Geographical coordinates	Number of soil samples	Number of yeast isolates	Number of cryptococcal isolates	Soil pH
West	Yanbu	24.0232° N, 38.1900° E	197	79	64	6.8
West	Jeddah	21.2854° N, 39.2376° E	100	29	12	6.8
Northwest	Umluj	25.0500° N, 37.2651° E	129	1	0	6.4
West	Medina	24.5247° N, 39.5692° E	50	2	0	6.4
East	Dammam	26.3927° N, 49.9777° E	56	2	0	7.00
Southwest	Alqunfudah	19.1281° N, 41.0787° E	32	1	0	7.00
Total			**562**	**114**	**76**	

Species identification: Total genomic DNA was extracted from all yeast-like colonies according to the protocol described in Xu, Vilgalys and Mitchell (2000). Species identification was achieved by sequencing the fungal barcoding region, internal transcribed spacer (ITS), encompassing both ITS1 and ITS2, and comparing it to available sequences on GenBank (Table S1). Genotypes and phenotypes of isolates identified as *C. deneoformans* through ITS sequences were further investigated.

Multilocus sequence typing (MLST): MLST sequences were obtained for all *C. deneoformans* isolates according to the ISHAM consensus scheme (Meyer et al. 2009). Alternate primers were designed for *LAC1*, *GPD1*, and *SOD1* based on the JEC21 whole genome sequence available on NCBI (https://www.ncbi.nlm.nih.gov/genome/?term=jec21) to ensure efficient amplification (Table S1). Primers for *CAP59*, IGS1, *PLB1* and *URA5* were unchanged from the consensus scheme. Sequences were manually edited using the software FinchTV V1.4.0 and aligned using the ClustalW algorithm in CLC sequence viewer V8.0. Allele types and sequence types (ST) were identified by making comparisons to sequences in the cryptococcal MLST Database (http://mlst.mycologylab.org/).

Serotype, molecular type and mating type: The serotypes and molecular types of the cryptococcal isolates were determined based on their MLST allele types and STs. Mating type was determined by primer-specific amplification of the *STE20α* and *STE20a* genes (Yan, Li, and Xu 2002).

Genetic diversity: Measures of genetic diversity within the *C. deneoformans* population were calculated using the software DnaSP V6: these included the total number of haplotypes (h), haplotype diversity (Hd), number of polymorphic sites (S), nucleotide diversity (π) which is the average number of nucleotide differences per site between two sequences, the average number of nucleotide differences (k) between any two randomly drawn sequences from the population and the Watterson estimator (θ_S).

Linkage disequilibrium and recombination: The index of association (I_A), a measure of linkage between loci in the population, and rbarD ($\bar{r}_d$), a less biased method that takes the number of sampled loci into account were calculated using the software Multilocus V1.3b (Agapow and Burt 2001). The observed value was compared to those based on randomized populations assuming infinite sexual recombination using 1,000 iterations to test the null hypothesis of complete linkage equilibrium (null hypothesis was rejected if $p < 0.05$). As another measure of recombination, we conducted the four-gamete test between each pair of MLST loci. In the simplest case of two loci with two alleles each, a maximum of four haplotypes are possible. The occurrence of no more than 3 of these haplotypes can be explained by mutation and the two loci are said to be phylogenetically compatible. If all four combinations are observed within the population, the two loci are said to be phylogenetically incompatible and assumed to be due to sexual recombination. For each pairwise comparison, graphs were generated where alleles of the two loci were listed side by side with lines connecting each observed combination of alleles. If the lines created an “hourglass” shape, this is indicative of the existence of all four possible haplotypes and thus phylogenetic incompatibility. Finally, the minimum number of recombination events that have occurred in the population’s evolutionary history was calculated using DnaSP.

Phylogenetic trees and goeBurst: To trace the possible origins of the Saudi Arabian *C. deneoformans* population, STs that share all or some alleles with Saudi Arabian

isolates were identified through the Fungal MLST database. Information pertaining to strains that belong to these STs were collected through a literature search on PubMed using the keywords "*Cryptococcus neoformans*" and "MLST" in all fields. Concatenated MLST sequences of the STs for which strain information could be identified were aligned to the Saudi Arabian STs in CLC sequence viewer. Neighbour-joining phylogenetic trees for concatenated sequences as well as for individual loci were constructed in MEGA7. A clustering analysis was conducted using the PHYLOViZ V2.0 software that utilizes the goeBURST algorithm to infer phylogenetic relationships between the STs and to determine the most likely founding genotype within clonal complexes (CC).

Antifungal resistance and melanin production: Fluconazole is the most commonly used first-line antifungal drug in treating primary and systemic cryptococcal infections in both immunocompromised and immunocompetent hosts (Perfect et al. 2010). The minimum inhibitory concentration (MIC) of fluconazole for the cryptococcal isolates was determined according to the protocol outlined by the Clinical and Laboratory Standards Institute (Rex et al. 2008) as a measure of their potential implications on clinical infection and management. Melanin production, an essential virulence factor, of the isolates was also quantified using spot densitometry on solid caffeic-acid agar medium following the protocol in Vogan *et al.* (2016).

3.5 Results

Yeast isolation from soil samples: The 562 soil samples yielded a total of 114 yeast isolates, 76 of which were identified as *C. deneoformans* based on ITS sequences (Table 1). The majority of cryptococcal strains (n = 64) were recovered from Yanbu while the rest (n = 12) were from Jeddah, both of which are located along the west coast of Saudi Arabia (Figure 1). Cryptococcal isolates accounted for 90% and 69% of all yeast isolates found in these two locations respectively. The other four sampled locations failed to yield any cryptococcal isolates.

Serotype, molecular type, mating type: MLST sequences (discussed below) revealed all 76 cryptococcal isolates to be serotype D, molecular type VNIV, thus identifying them as strains of *C. deneoformans.* Mating type-specific amplification of

STE20 gene revealed 75 isolates to be mating-type (*MAT*)**α** while one isolate was *MAT***a**.

Table 2: The list of *C. deneoformans* isolates found in Saudi Arabia. Their MLST genotypes, mating types and MIC to fluconazole are listed here.

Strain ID	City	*CAP59*	*GPD1*	*IGS1*	*LAC1*	*PLB1*	*SOD1*	*URA5*	ST	*MAT*	MIC (Fluc.)
J1	Jeddah	16	21	30	19	13	17	19	160	α	2
J100	Jeddah	16	21	30	19	13	17	19	160	α	2
J18	Jeddah	16	21	30	19	13	17	19	160	α	2
J20	Jeddah	16	21	30	19	13	17	19	160	α	2
J26	Jeddah	16	21	30	19	13	17	19	160	α	1
J4	Jeddah	16	21	30	19	13	17	19	160	α	1
J6	Jeddah	16	21	30	19	13	17	19	160	α	2
J62	Jeddah	16	21	30	19	13	17	19	160	α	2
J7	Jeddah	16	21	30	19	13	17	19	160	α	4
J82	Jeddah	16	21	30	19	13	17	19	160	α	2
J85	Jeddah	16	21	30	19	13	17	19	160	α	2
J86	Jeddah	16	21	30	19	13	17	19	160	α	2
Y101	Yanbu	16	21	30	19	13	17	19	160	α	2
Y102	Yanbu	16	21	30	19	13	17	19	160	α	2
Y103	Yanbu	16	21	30	19	13	17	19	160	α	2
Y104	Yanbu	16	21	30	19	13	17	19	160	α	2
Y105	Yanbu	16	21	30	19	13	17	19	160	α	2
Y106	Yanbu	16	21	30	19	13	17	19	160	α	1
Y111	Yanbu	16	21	30	19	13	17	19	160	α	2
Y113	Yanbu	16	21	30	19	13	17	19	160	α	1
Y114	Yanbu	16	21	30	19	13	17	19	160	α	2
Y115	Yanbu	16	21	30	19	13	17	19	160	α	2
Y116	Yanbu	16	21	30	19	13	17	19	160	α	1
Y121	Yanbu	16	21	30	19	13	17	19	160	α	1
Y122	Yanbu	16	21	30	19	13	17	19	160	α	2
Y124	Yanbu	16	21	30	19	13	17	19	160	α	2
Y126	Yanbu	16	21	30	19	13	17	19	160	α	2
Y128	Yanbu	16	21	30	19	13	17	19	160	α	2
Y132	Yanbu	16	21	30	19	13	17	19	160	α	2
Y134	Yanbu	16	21	30	19	13	17	19	160	α	2
Y137	Yanbu	16	21	30	19	13	17	19	160	α	2
Y140	Yanbu	16	21	30	19	13	17	19	160	α	2
Y141	Yanbu	16	21	30	19	13	17	19	160	α	2
Y142	Yanbu	16	21	30	19	13	17	19	160	α	2
Y143	Yanbu	16	21	30	19	13	17	19	160	α	2

Y144	Yanbu	16	21	30	19	13	17	19	160	α	2
Y145	Yanbu	16	21	30	19	13	17	19	160	α	1
Y147	Yanbu	16	21	30	19	13	17	19	160	α	2
Y152	Yanbu	16	21	30	19	13	17	19	160	α	2
Y153	Yanbu	16	21	30	19	13	17	19	160	α	2
Y156	Yanbu	16	21	30	19	13	17	19	160	α	1
Y158	Yanbu	16	21	30	19	13	17	19	160	α	4
Y161	Yanbu	16	21	30	19	13	17	19	160	α	1
Y162	Yanbu	16	21	30	19	13	17	19	160	α	2
Y163	Yanbu	16	21	30	19	13	17	19	160	α	2
Y165	Yanbu	16	21	30	19	13	17	19	160	α	2
Y170	Yanbu	16	21	30	19	13	17	19	160	α	1
Y171	Yanbu	16	21	30	19	13	17	19	160	α	1
Y172	Yanbu	16	21	30	19	13	17	19	160	α	2
Y176	Yanbu	16	21	30	19	13	17	19	160	α	2
Y177	Yanbu	16	21	30	19	13	17	19	160	α	1
Y178	Yanbu	16	21	30	19	13	17	19	160	α	2
Y182	Yanbu	16	21	30	19	13	17	19	160	α	2
Y184	Yanbu	16	21	30	19	13	17	19	160	α	2
Y189	Yanbu	16	21	30	19	13	17	19	160	α	2
Y190	Yanbu	16	21	30	19	13	17	19	160	α	2
Y192	Yanbu	16	21	30	19	13	17	19	160	α	2
Y194	Yanbu	16	21	30	19	13	17	19	160	α	2
Y195	Yanbu	16	21	30	19	13	17	19	160	α	1
Y196	Yanbu	16	21	30	19	13	17	19	160	α	1
Y30	Yanbu	16	21	30	19	13	17	19	160	α	1
Y67	Yanbu	16	21	30	19	13	17	19	160	α	1
Y9	Yanbu	16	21	30	19	13	17	19	160	α	1
Y107	Yanbu	16	21	24	20	13	22	32	294	α	1
Y108	Yanbu	16	21	24	20	13	22	32	294	α	1
Y119	Yanbu	16	21	24	20	13	22	32	294	α	1
Y130	Yanbu	16	21	24	20	13	22	32	294	α	1
Y133	Yanbu	16	21	24	20	13	22	32	294	α	1
Y138	Yanbu	16	21	24	20	13	22	32	294	α	1
Y157	Yanbu	16	21	24	20	13	22	32	294	α	1
Y164	Yanbu	16	21	24	20	13	22	32	294	α	1
Y187	Yanbu	16	21	24	20	13	22	32	294	α	1
Y38	Yanbu	26	21	30	22	13	19	18	499	a	1
Y112	Yanbu	16	21	24	19	13	22	32	613	α	2
Y97	Yanbu	16	46	24	20	13	20	32	614	α	1
Y87	Yanbu	24	21	91	21	14	17	34	615	α	2

MLST and genetic diversity: MLST sequences were obtained for all 76 *C. deneoformans* isolates (Table 2). The concatenated sequences of the 7 loci added up to a total length of 3,960bp. The 76 isolates belonged to 6 STs. ST160 was the most frequent (n = 63), followed by ST294 (n = 9). The remaining four STs, ST499, ST613, ST614 and ST615, were singletons with the latter three being novel. With an Hd of 0.295, *URA5* was the most polymorphic locus in our population with 4 haplotypes and 14 single-nucleotide polymorphisms (SNP) (Table 3). *LAC1* and *SOD1* also contained 4 haplotypes each with 7 and 4 SNPs respectively. *GPD1* and *PLB1* were the least diverse with all isolates but one sharing the same alleles at the two loci. Overall, the population showed low genetic diversity with an Hd of 0.302 and π of 0.0013 for the concatenated MLST sequences (Table 3).

Table 3: Genetic diversity in Saudi Arabian *C. deneoformans* population. Haplotype diversity (Hd), nucleotide diversity (π), average number of nucleotide differences (k) and Watterson's estimator (θS) were calculated using DnaSP.

Locus	Length (bp)	Polymorphic sites (S)	Haplotypes (h)	Haplotype diversity (Hd)	Nucleotide diversity (π)	Average number of nucleotide difference (k)	Theta per site (θ_S)
CAP59	560	2	3	0.052	0.00009	0.053	0.00073
GPD1	546	1	2	0.026	0.00005	0.026	0.00038
IGS1	697	5	3	0.273	0.00054	0.375	0.00147
LAC1	473	7	4	0.277	0.00267	1.262	0.00302
PLB1	517	1	2	0.026	0.00005	0.026	0.00039
SOD1	528	4	4	0.277	0.00105	0.554	0.00155
URA5	639	14	4	0.295	0.00456	2.883	0.00452
All loci	**3960**	**34**	**6**	**0.302**	**0.00131**	**5.179**	**0.00176**

Clonality and recombination: The over-representation of two multilocus genotypes suggests that the Saudi Arabian *C. deneoformans* population has a significant clonal component. To examine whether there is any evidence of recombination, I_A and $\bar{r}_d$ of the *C. deneoformans* population were measured both before and after clonally correcting the dataset. In the former, I_A and $\bar{r}_d$ were 2.98 and 0.59 respectively whereas the values were lower (1.08 and 0.18 respectively) in the clone-corrected dataset. In both cases, the null hypothesis of linkage equilibrium and sexual recombination was rejected at $p<0.001$ and $p<0.05$ respectively, suggesting that the population is highly clonal. Four-gamete tests were conducted between each pair of MLST loci adding up to a total of 21 pairwise comparisons. Each observed allelic combination was visualized by connecting the alleles with lines: none of the 21 generated graphs contained the "hourglass" shape that would be indicative of

recombination between loci (five of these graphs are shows in Figure 2). However, DnaSP detected four possible recombination events among the SNPs in the concatenated MLST sequences of the Saudi Arabian STs. Furthermore, ST613 could be explained as a recombinant product between the two major Saudi Arabian STs, ST160 and ST294 (Figure 3). Specifically, ST613 is identical to ST160 at the *LAC1* locus while it shares IGS1, *SOD1* and *URA5* alleles with ST294. Alleles at *CAP59, GPD1,* and *PLB1* loci for ST613 could have been inherited from either STs as ST160 and ST294 are identical at these three loci.

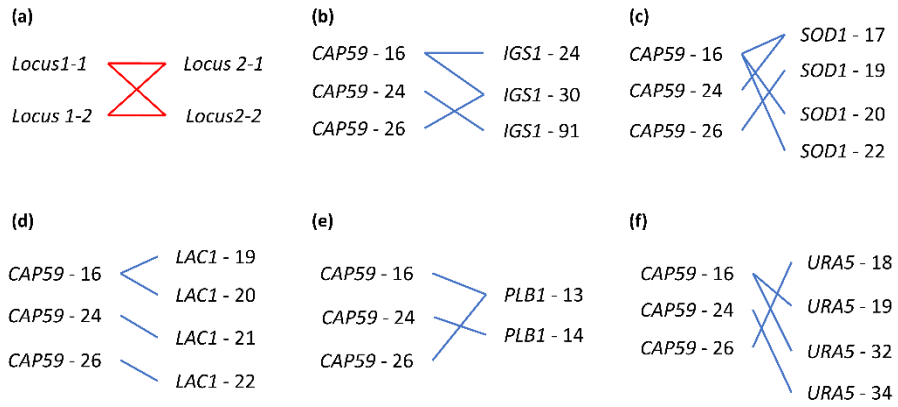


Figure 2: Test of phylogenetic incompatibility using the four-gamete test. (A) Two loci with two alleles each are said to be phylogenetically incompatible if all four possible allelic combinations are observed. This is visualized by the formation of an 'hourglass' shape when alleles are connected by lines. (B)-(F) Five of the 21 graphs generated from the four-gamete tests are shown here. The observed allelic combinations in the Saudi Arabian *C. deneoformans* population are shown by connecting lines. None of these graphs show the hourglass shape and thus fail to show indications of sexual recombination.

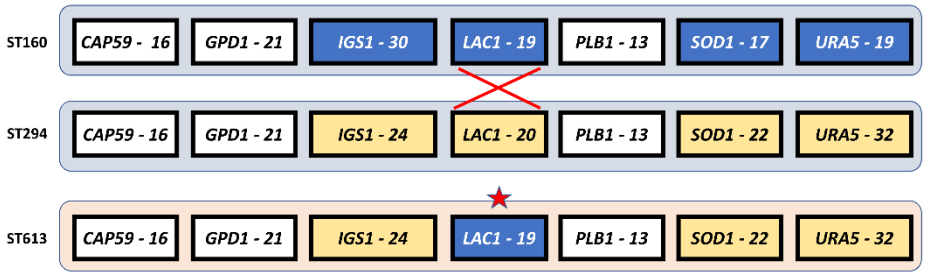


Figure 3: Sexual recombination in the Saudi Arabian *C. deneoformans* population. The newly identified ST, ST613, seems to be a product of sexual recombination between the two most frequent STs, ST160 and ST294, in the Saudi Arabian cryptococcal population. IGS1, SOD1 and URA5 alleles in ST613 are identical to that of ST294 while the LAC1 allele is identical to that of ST160. CAP59, GPD1 and PLB1 could have been derived from either ST.

Comparison with published MLST data: We identified 100 STs in the cryptococcal MLST database that shared at least one allele with our Saudi Arabian STs. From the literature search on PubMed, we were only able to obtain strain information for 56 of these 100 STs. Thus, only these 56 STs were included in our comparative. In total, we found 157 strains belonging to the 56 STs; the vast majority were of clinical (human or veterinary) origin (n=146) while only 11 strains had been isolated from the environment (Table 4). ST160, the dominant genotype in Saudi Arabia, has been previously found in clinical isolates from the United States (n=5), France (n=3), Canada (n=2), Germany (n=1), Columbia (n=1), Brazil (n=1) and a veterinary isolate from Italy (n=1) (Andrade-Silva et al. 2018; Massimo Cogliati, Zani, et al. 2016; Danesi et al. 2014; Desnos-Ollivier et al. 2015; Himeshi Samarasinghe et al. 2018). ST294 has been found in 2 veterinary isolates from Italy (Danesi et al. 2014). ST499 has been previously reported in environmental isolates from Greece (n=6) (Massimo Cogliati, Zani, et al. 2016). A complete minimum spanning tree of the Saudi Arabian STs and related global STs was generated using the goeBURST algorithm in PHYLOViZ. ST121, a genotype found exclusively in France, was determined to be the most likely founding genotype with six major clusters and two singletons branching off from it (Figure 4). ST160 is derived from ST121 and is a sub-group founder itself. ST499 is loosely derived from ST160 at a 3-degree separation. ST615 is another subgroup founder with several global STs descended from it. ST294 is directly descended from ST613 while ST614 is derived from ST294. Clustering by geographical origin failed to reveal any patterns. Four-gamete tests were conducted between all 21 MLST loci pairs in the global ST dataset: the hourglass shape was

observed in each of the 21 pairwise comparisons, consistent with frequent recombination in the global sample (an example is shown in Figure 5a). In fact, ST160 can be constructed by pairing several other global STs. For an example, ST160 could be potentially derived from recombination between ST117 and ST118 (Figure 5b).

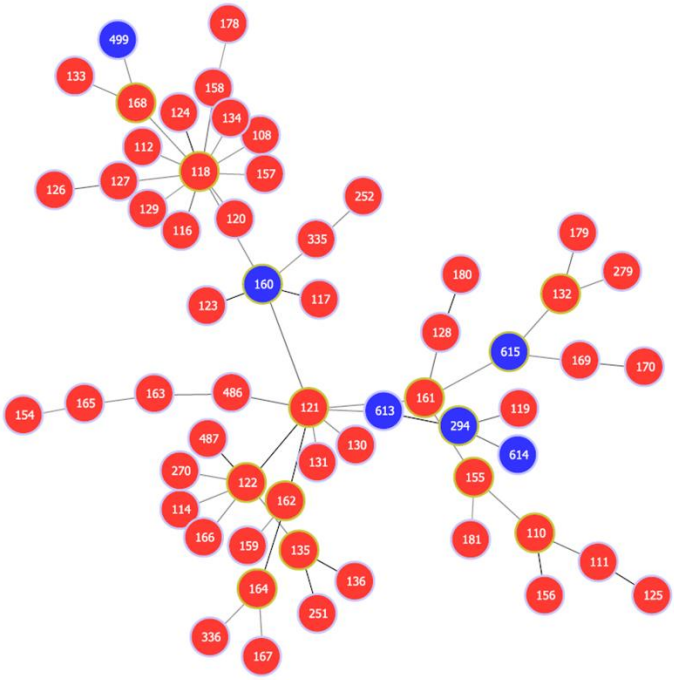


Figure 4: Inferred phylogenetic relationships among global *C. deneoformans* STs. This minimum spanning tree was generated in PHYLOViZ via the goeBURST algorithm. The Saudi Arabian STs are highlighted in blue. ST121, found in a French clinical strain, is the most likely founding genotype. ST160 and ST615 are subgroup founders while ST499, ST613, ST614 and ST294 appear to be more recently derived.

Table 4: Sequence types (STs) that share alleles with Saudi Arabian STs and associated strain information in the published literature.

ST	strains	Country	Source	Reference	loci	*CAP59*	*GPD1*	*IGS1*	*LAC1*	*PLB1*	*SOD1*	*URA5*	type
160	5	USA	Cln	f	7	16	21	30	19	13	17	19	VNIV
	3	France	Cln	b									
	2	Canada	Cln	f									
	1	Italy	Vet	c									
	1	Germany	Cln	a									
	1	Colombia	Cln	a									
	1	Brazil	Cln	g									
294	2	Italy	Vet	c	7	16	21	24	20	13	22	32	VNIV
499	6	Greece	Env	a	7	26	21	30	22	13	19	18	VNIV
117	1	Australia	Cln	a	6	16	21	30	19	13	1	19	VNIV
123	1	France	Cln	b	6	16	21	43	19	13	17	19	VNIV
168	1	Germany	Cln	a	6	22	21	30	22	14	17	18	VNIV
	8	Japan	Cln	a									
	3	France	Cln	b									
118	1	France	Cln	b	5	16	21	30	13	14	17	16	VNIV
119	1	France	Cln	b	5	16	21	31	20	13	20	32	VNIV
121	2	France	Cln	a	5	16	21	32	19	13	17	20	VNIV
	17	France	Cln	b									
130	1	France	Cln	b	5	22	21	32	19	13	22	18	VNIV
131	1	France	Cln	b	5	22	21	32	19	14	17	19	VNIV
133	2	France	Cln	b	5	26	21	30	22	14	21	22	VNIV
158	1	France	Cln	b	5	16	21	24	13	14	17	36	VNIV
161	1	France	Cln	b	5	16	21	32	21	14	17	21	VNIV
169	1	France	Cln	b	5	24	21	26	21	14	20	32	VNIV
170	1	France	Cln	b	5	24	21	26	21	14	32	18	VNIV
110	1	Germany	Cln	a	4	15	21	24	21	13	20	22	VNIV
	2	France	Cln	b									
116	7	Germany	Not known (4), Cln (3)	a	4	16	21	29	13	14	17	24	VNIV
	4	France	Cln	b									
120	1	France	Cln	b	4	16	21	31	13	14	19	20	VNIV
122	1	France	Cln	a	4	16	21	32	24	13	17	20	VNIV
	9	France	Cln	b									
124	1	France	Cln	b	4	16	21	45	13	14	17	16	VNIV
125	1	France	Cln	a	4	16	21	45	21	13	21	22	VNIV
	4	France	Cln	b									
132	1	France	Cln	b	4	25	21	31	15	14	17	34	VNIV
155	1	France	Cln	b	4	15	21	52	21	14	17	22	VNIV
156	1	France	Cln	b	4	15	21	24	21	13	20	35	VNIV
157	1	France	Cln	b	4	16	20	30	22	14	30	17	VNIV
162	1	France	Cln	b	4	16	21	32	19	20	17	20	VNIV
163	1	France	Cln	b	4	16	22	32	19	13	23	18	VNIV
179	1	France	Cln	b	4	25	21	53	15	13	17	34	VNIV
279	1	Italy	Cln	a	4	22	22	31	22	14	17	34	VNIV
335	1	Colombia	Cln	a	4	27	28	30	19	14	17	41	VNIV

487	1	Germany	Cln	a	4	16	21	32	24	13	17	32	VNIV
108	1	France	Cln	b	3	14	21	42	13	14	18	34	VNIV
111	1	France	Cln	b	3	15	21	45	21	13	21	22	VNIV
112	2	Italy	Cln, Vet	a	3	16	22	31	24	14	17	16	VNIV
	4	France	Cln	b									
	1	Italy	Vet	c									
114	8	France	Cln	b	3	16	21	26	24	18	17	22	VNIV
	3	USA	Cln	f									
126	1	Thailand	Env	a, d	3	17	21	28	19	14	1	20	VNIV
128	1	France	Cln	b	3	20	21	32	21	19	17	21	VNIV
134	1	France	Cln	b	3	26	24	33	23	14	17	16	VNIV
136	1	France	Cln	b	3	27	22	30	24	13	17	20	VNIV
159	1	France	Cln	b	3	16	21	24	16	20	20	16	VNIV
164	1	France	Cln	b	3	16	22	32	19	20	17	20	VNIV
166	1	France	Cln	b	3	16	24	26	24	13	17	31	VNIV
178	1	France	Cln	b	3	29	24	44	22	14	14	36	VNIV
180	1	Denmark	Env	a	3	20	21	26	21	19	17	21	VNIV
	1	France	Cln										
	2	France	Cln	b									
	3	Thailand	Env	f									
251	1	Italy	Vet	c	3	26	22	43	24	13	17	20	VNIV
252	1	Italy	Vet	a	3	22	22	30	19	14	23	41	VNIV
	1	Italy	Vet	c									
270	1	Thailand	Env	f	3	16	28	53	24	14	17	20	VNIV
486	1	Germany	Cln	a	3	16	22	32	19	13	17	22	VNIV
127	1	Korea	Cln	e	2	17	21	28	13	14	1	16	VNIV
129	1	France	Cln	b	2	21	21	44	17	14	23	23	VNIV
135	1	France	Cln	a	2	27	22	43	24	13	17	20	VNIV
	6	Italy	Cln (5), Vet (1)	a									
	1	Virginia, USA	Env	a									
	6	France	Cln	b									
	1	Italy	Vet	c									
	1	USA	Cln	f									
165	2	France	Cln	b	2	16	22	48	24	13	23	17	VNIV
	2	USA	Cln	f									
181	1	France	Cln	b	2	32	21	52	21	22	31	16	VNIV
336	2	Columbia	Cln	a	2	16	22	32	14	14	18	17	VNIV
154	1	France	Cln	b	1	14	22	67	13	14	23	17	VNIV
167	1	France	Cln	b	1	17	22	32	14	21	17	20	VNIV

Cln = clinical, Env = environmental, Vet = veterinary.

[a](Cogliati *et al.* 2016), [b](Desnos-Ollivier *et al.* 2015), [c](Danesi *et al.* 2014), [d](Kaocharoen *et al.* 2013), [e](Park *et al.* 2015), [f](Samarasinghe *et al.* 2018a), [g](Andrade-Silva *et al.* 2018)

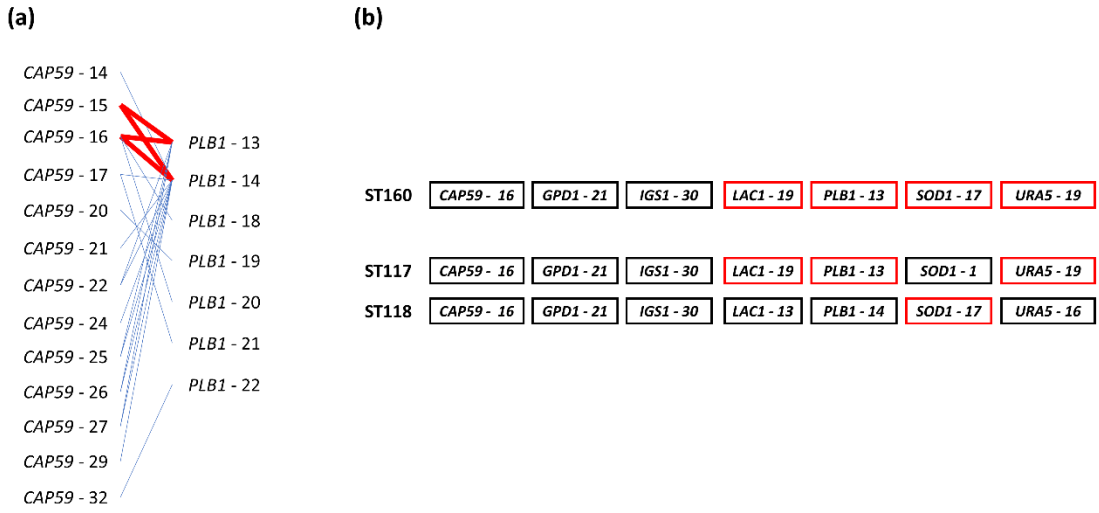


Figure 5: Evidence for recombination in the global *C. deneoformans* population. (A) The graph generated by the four-gamete test between CAP59 and PLB1 in the global *C. deneoformans* STs. Hourglass shape indicates sexual recombination occurring between the two loci. (B) ST160 can be derived from recombination between ST117 and ST118. ST160 shares SOD1 with ST118 while sharing LAC1, PLB1 and URA5 with ST117. CAP59, GPD1 and IGS1 could have been inherited from either ST.

Phylogenetic trees: The neighbour-joining tree created using concatenated MLST sequences of the 6 Saudi Arabian STs and the 56 partially related STs is shown in Figure 6. The STs are distributed into five major clusters. ST294, along with ST614 and ST613, appear to be three of the four most-recently derived STs in the analyzed population. All three are closely related to each other. In contrast, ST160, ST499 and ST615 are located in three separate clades and are genetically more similar to non-Saudi Arabian STs. For an example, the closest relative of ST160 is ST123, an ST found in a French clinical strain. We next created individual neighbour-joining trees for the 7 MLST loci to trace possible origins of the Saudi Arabian alleles. To avoid redundancy, a single, randomly picked representative was used for STs that shared alleles at a particular locus. *GPD1*, IGS1 and *PLB1* alleles found in the Saudi Arabian *C. deneoformans* population seem to be more recently derived as indicated by their locations at the tips of the shortest branches in the phylogenetic trees (Figure 7). *LAC1* alleles from ST294 and ST614 appear to be recently derived while those of ST160, ST499, ST613 and ST615 appear to have a more ancient origin. The Saudi Arabian *CAP59*, *SOD1* and *URA5* alleles do not follow this pattern and are dispersed throughout their respective trees.

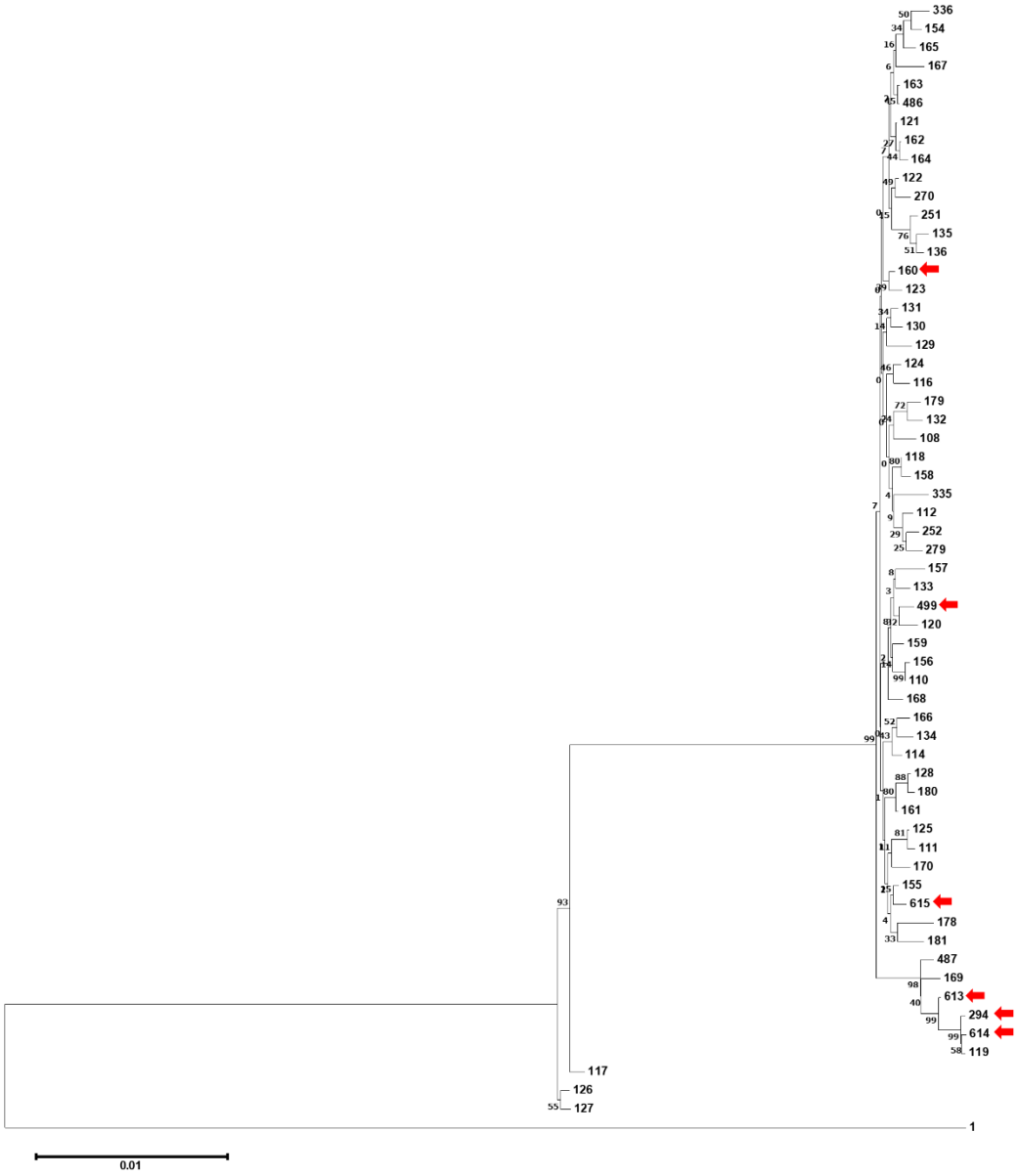


Figure 6: Neighbour-joining tree of concatenated MLST sequences of Saudi Arabian STs and related global STs. The Saudi Arabian STs are indicated by the red arrows.

Antifungal resistance and melanin production: All 76 Saudi Arabian *C. deneoformans* isolates were susceptible to the antifungal drug fluconazole with the highest MIC recorded being 4µg/mL (n=2). 46 isolates had an MIC of 2µg/mL while the rest (n=28) had an MIC of 1µg/mL. Melanin production varied significantly among isolates, ranging from no observable melanin production to high melanin production following 7 days of incubation at 30°C (Figure 7a). Average melanin

production differed significantly between strains of the two major sequence types, ST160 and ST294 ($p<0.001$, Figure 7b). Melanin production was not significantly correlated with the sampling location of the isolates (Figure 8).

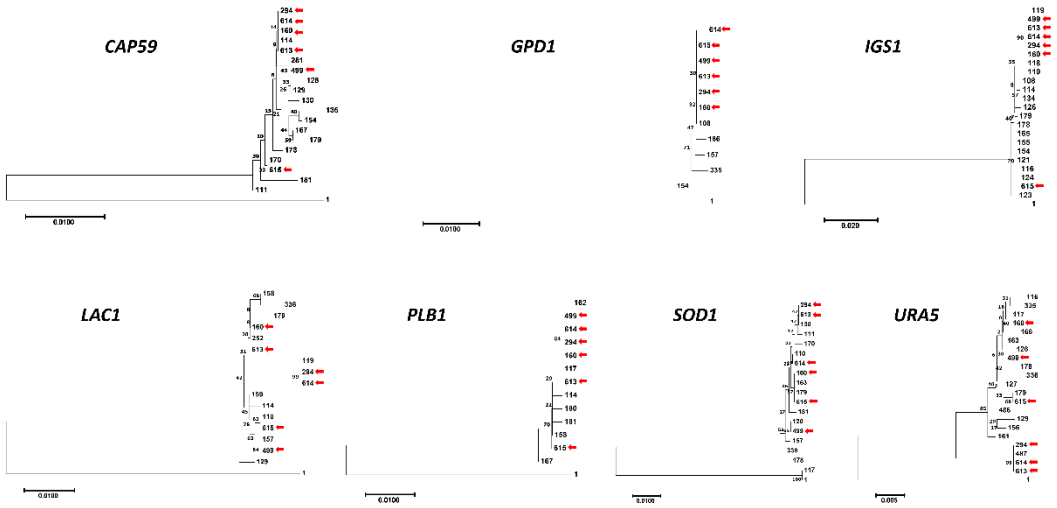


Figure 7: Neighbour-joining trees of the seven MLST loci. Alleles from the six Saudi Arabian STs and representatives of other global STs were used to construct neighbour-joining trees for each of the seven MLST loci. Saudi Arabian STs are indicated by the red arrows.

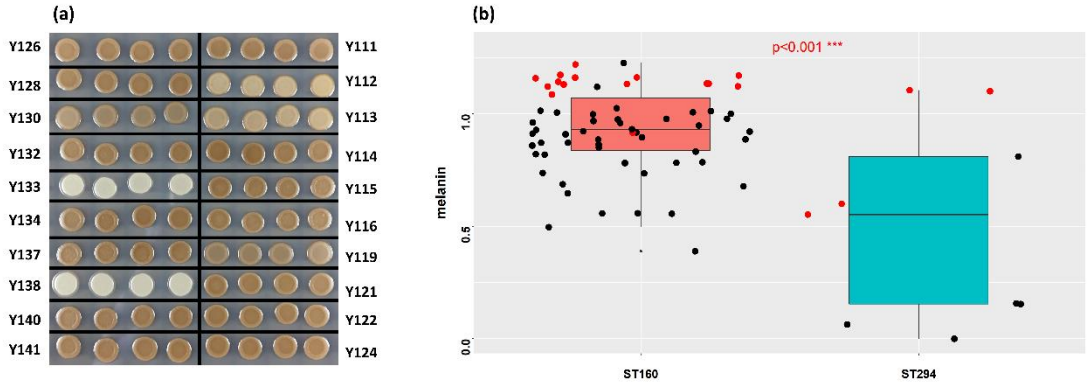


Figure 8: Melanin production of the Saudi Arabian *C. deneoformans* isolates. (A) Melanin quantification assay of 20 selected isolates from the Saudi Arabian population. (B) Melanin production of the isolates grouped by MLST sequence type. Isolates shown in panel (A) are highlighted in red. ST160 isolates on average produced significantly more melanin than ST294 isolates ($P < 0.01$).

3.6 Discussion

In what was initially a routine survey of yeast presence in Saudi Arabian soil, we unexpectedly discovered a large *C. deneoformans* population which turned out to be the predominant yeast species in the sampled locations. This discovery reveals a novel ecological niche for *C. deneoformans*, whose preferred habitat was previously believed to be subtropical and temperate climates. Our analyses suggest that a recent introduction, possibly through human migration, followed by expansion via clonal reproduction and limited sexual recombination has led to the establishment of a largely clonal *C. deneoformans* population in Saudi Arabia with several, novel genotypes. The presence of both ancestral and newly derived genotypes indicates that Saudi Arabia could potentially act as both a source and a sink for *C. deneoformans* genotypes in a global context.

C. deneoformans (VNIV) isolates of clinical and environmental origins have been recovered from all continents except Africa and Antarctica. However, strains of other molecular types (VNI, VNII, VNIII, *C. gattii*) often outnumber VNIV strains even in Europe, the largest natural reservoir of *C. deneoformans* where 30% of all cryptococcal infections are caused by VNIV strains (Massimo Cogliati 2013). Our study describes the first natural cryptococcal population identified to date that is made up exclusively of VNIV strains. Furthermore, our results point to a previously unidentified ecological niche for *C. deneoformans*. Yanbu has a desert climate with an annual precipitation of 91mm and an average temperature of 26.5°C: during summer

months the mean daily temperature can reach mid 30°Cs. Jeddah, likewise, has a desert climate with temperatures reaching mid 40°C during summer months, and an annual precipitation of 52mm (Almazroui et al. 2012; Krishna 2014). *C. deneoformans* was previously thought to be less thermotolerant than *C. neoformans* and is typically associated with temperate climates. In addition to being the first *C. deneoformans* population to be discovered in a desert climate, the Saudi Arabian population appears to be well-established and thriving in this novel niche.

The sporadic appearance of cryptococcal infections in Saudi Arabia over the past few decades suggests that *C. deneoformans* may have persisted in the environment for a significant period of time. Cases of cryptococcal infections in both immunocompromised and immunocompetent individuals in Saudi Arabia have been reported as far back as 1989 (Al-Hedaithy 1992). In most cases, serotype/molecular type of the isolates were not determined (Al-Tawfiq and Ghandour 2007; Alsum et al. 2012; Mansoor et al. 2015). However, in 2005, an apparently healthy male with primary cutaneous cryptococcosis was determined to be infected with a *C. deneoformans* isolate (Al-Marzooq et al. 2005). *C. neoformans* was first recovered from environmental sources in Saudi Arabia in 2015 during a survey of pigeon droppings: *C. neoformans* was identified based on melanin synthesis, ability to grow at 37°C, and urease production (Abulreesh et al. 2015). Serotype and molecular type of these isolates were not determined. Given the geographical proximity of their sampling locations in Mecca to Yanbu (382km) and Jeddah (86km), it is possible that some of these isolates are *C. deneoformans*. In this survey, we failed to identify any cryptococcal isolates in other four sampled locations. Temperature is likely not a contributing factor as Jeddah and Mecca are two of the driest and warmest cities in Saudi Arabia. Furthermore, we cannot explain the absence of *C. neoformans* in Saudi Arabian soil.

Our phylogenetic analyses suggest that some of *C. deneoformans* genotypes may be introduced into Saudi Arabia via anthropogenic activities, including human travel. Due to its thriving economy centered around the oil industry, Saudi Arabia has become one of the top five destinations for international migrant workers (United Nations 2017). As of 2017, 37% of its total population of 33 million is comprised of

international migrants, largely from other Asian and African countries (United Nations 2017). In addition, the annual Hajj and Umrah pilgrimages attract over 20 million domestic and foreign visitors to Mecca every year (General Authority for Statistics Kingdom of Saudi Arabia 2017b, 2017a). Frequent international travel between Saudi Arabia and other countries likely facilitated the introduction and exchange of *C. deneoformans*. ST160, the most frequent genotype found in the Saudi Arabian population, has a global presence with strains having been isolated from North America, South America and Europe. ST160 strains are likely adept at surviving a wide range of environmental conditions, facilitating its spread across continents. Significant genetic differences were observed between Saudi Arabian STs. For example, the two dominant STs, ST160 and ST294, were found in different clades on the concatenated phylogenetic tree (Figure 4). ST499 and ST615 were more closely related to other global STs than to other genotypes found in Saudi Arabia (Figure 4, Figure 6). The novel IGS1 allele recovered from ST615 is significantly different from that of ST160 (4 nucleotide insertions and 5 SNPs) and ST294 (4 nucleotide insertions and 4 SNPs). It is closely related to IGS1 of clinical strains from Europe and USA. In fact, ST615 differs from the rest of the population at 5 of the 7 MLST loci. Furthermore, significant phenotypic differences exist between Saudi Arabian STs as highlighted by differential melanin productions of strains both within and between ST160 and ST294 (Figure 7). These results point to multiple introductory events of genetically distinct *C. deneoformans* strains into Saudi Arabia with genotypes more suited for survival in the desert climate taking hold and persisting in the environment.

Saudi Arabia could serve as both a source and a sink for the *C. deneoformans* global population. ST160 and ST615 are founding genotypes for many other global isolates and their MLST alleles appear to be relatively ancestral in origin. On the other hand, ST294, ST614, ST613 and ST499 are more recently derived according to both goeBURST analysis and neighbour-joining phylogenetic trees. Frequent international travel into and out of Saudi Arabia likely facilitates the ingress and egress of *C. deneoformans* isolates in its environment. Mediterranean Europe is currently considered the point of origin for *C. deneoformans* due to the high prevalence and

genetic diversity of this species in this region. Alternatively, both the unique and broad distribution of the Saudi Arabian STs suggests that Middle East could also be the source of the contemporary global *C. deneoformans* population. In fact, Mecca is host to an unusually large population of pigeons, a well-known carrier of *Cryptococcus* that played a pivotal role in the historical expansion of cryptococcal species (Casadevall et al. 2017). It is possible that the true niche of *C. deneoformans* is the semi-arid deserts found in Africa and the Middle East, especially given that its sister species *C. neoformans* is believed to have originated on the African continent prior to global dispersal (Anastasia P. Litvintseva and Mitchell 2012). However, the low genetic diversity of the Saudi Arabian *C. deneoformans* population suggests that Saudi Arabia is more likely a contemporary source of novel genotypes as opposed to the point of ancestral origins.

Most strains from other geographic regions but that share alleles with Saudi Arabian population are of clinical origin (Table 3). This could be a result of sampling bias where clinical isolates were more often studied and genotyped than environmental strains. On the other hand, clinical strains, potentially with higher expression of virulence factors, can be genetically and phenotypically better suited for spread and survival in a variety of climates: their propagation is likely facilitated by their human hosts' travels. Indeed, cryptococcal infections can stay dormant inside patients for decades, increasing the likelihood of their introduction to new geographic regions (Dromer, Ronin, and Dupont 1992; Garcia-Hermoso, Janbon, and Dromer 1999). In 2007, a 32-year old immunocompetent Indonesian female living in Saudi Arabia was diagnosed with cryptococcal vertebral osteomyelitis; although she had not travelled to Indonesia within the 7 years prior to her diagnosis, the possibility of cryptococcal spores from her native Indonesia lying dormant within the body cannot be discounted (Al-Tawfiq and Ghandour 2007).

Saudi Arabian *C. deneoformans* population shows genetic signatures of both sexual and asexual reproduction. The presence of *MAT**a*** strains, along with ST613, a very likely recombinant product between the two major genotypes, support the occurrence of sexual recombination in the Saudi Arabian *C. deneoformans* population. Low prevalence of *MAT**a*** (1/76) is consistent with mating type ratios observed in *C.*

neoformans and *C. deneoformans* populations in other regions (Chowdhary et al. 2011; Feng et al. 2008; Freire et al. 2012). In fact, previous MLST sequencing of global *C. deneoformans* isolates revealed the population structure of this species to be recombinant, rather than clonal (Massimo Cogliati, Zani, et al. 2016). However, the low genetic diversity of the Saudi Arabian population suggests that asexual reproduction is still the main mode of expansion in this region. The novel *GPD1* allele in ST614 differs from that of the rest of the population by a single SNP, which is likely a result of mutation accumulated during mitotic replication.

In this study, we obtained 76 *C. deneoformans* isolates from Saudi Arabian soils through a culture-based method. While this method allowed us to selectively enrich yeasts for isolation, under certain circumstances such as the presence of other microbes capable of inhibiting/killing *C. deneoformans*, our method could produce an under-estimate of the prevalence of *C. deneoformans*. However, we would like to note that we extracted total DNA from 10 soil samples from Yanbu and Jeddah, with five representing soils that we successfully isolated *C. deneoformans* and five that did not. We then used *Cryptococcus* species-specific primers to amplify the extracted soil DNA. However, none of the 10 samples showed any amplification (data not shown). Our results suggest that even if existed, cryptococcal cells and DNA in soils were likely in very low concentration. Thus, to have accurate assessments of fungal diversity in the Saudi Arabian soils, including *C. deneoformans* prevalence and diversity, a combination of culture-dependent and culture-independent methods would be needed.

In conclusion, the Saudi Arabian *C. deneoformans* population obtained here represents an expansion of this species' previously known ecological niche to include desert climates. Although largely clonal, the population carries several alleles and STs that are unique to Saudi Arabia, highlighting the combined role of asexual and sexual reproduction in its propagation. Over time, the Saudi Arabian *C. deneoformans* population will likely establish its own unique genetic signature through mutation and recombination, especially given its extreme environmental conditions compared to previously known, natural *C. deneoformans* habitats.

3.7 Funding

This work was supported by a graduate scholarship from the Cultural Bureau of Saudi Arabia (to R.A.) and by grants from the Natural Sciences and Engineering Research Council (NSERC Discovery Grant 531998 and CRDPJ 474638 – 14) of Canada.

3.8 Acknowledgements

We would like to thank the family of Renad Aljohani for their assistance in obtaining the soil samples from Saudi Arabia.

3.9 Supplementary Data

Table S1: Primer sets used in this study.

Primer	Sequence (5' to 3')
ITS1	TCCGTAGGTGAACCTGCGG
ITS4	TCCTCCGCTTATTGATATGC
GPD1F	ATTCAACGACGTACTCGGCA
GPD1R	GGTTGGAATCAACGGTTTCGG
LAC1F	GGAGTGGCTAGAGCTGCAAT
LAC1R	GCACAATTGCAGCAACGCTT
SOD1F	TGTCAACCATAGCGACACCG
SOD1R	TCCCACCAAATACGCCTCAA

3.10 References

Abulreesh HH, Organji SR, Elbanna K *et al.* First report of environmental isolation of *Cryptococcus neoformans* and other fungi from pigeon droppings in Makkah, Saudi Arabia and in vitro susceptibility testing. *Asian Pacific J Trop Dis* 2015;**5**:622–6.

Agapow PM, Burt A. Indices of multilocus linkage disequilibrium. *Mol Ecol Notes* 2001;**1**:101–2.

Al-Hedaithy SSA. The medically important yeasts present in clinical specimens. *Ann Saudi Med* 1992;**12**:57–62.

Al-Marzooq YM, Chopra R, Al-Mommatten MI *et al.* Fine-needle aspiration diagnosis of primary cutaneous cryptococcosis in an immunocompetent patient: A case report. *Diagn Cytopathol* 2005;**32**:219–21.

Al-Tawfiq JA, Ghandour J. *Cryptococcus neoformans* abscess and osteomyelitis in an immunocompetent patient with tuberculous lymphadenitis. *Infection* 2007;**35**:377–82.

Almazroui M, Nazrul Islam M, Athar H *et al.* Recent climate change in the Arabian Peninsula: annual rainfall and temperature analysis of Saudi Arabia for 1978-2009. *Int J Climatol* 2012;**32**:953–66.

Alsum Z, Al-Saud B, Al-Ghonaium A *et al.* Disseminated cryptococcal infection in patient with novel JAK3 mutation severe combined immunodeficiency, with resolution after stem cell transplantation. *Pediatr Infect Dis J* 2012;**31**:204–6.

Andrade-Silva LE, Ferreira-Paim K, Ferreira TB *et al.* Genotypic analysis of clinical and environmental *Cryptococcus neoformans* isolates from Brazil reveals the presence of VNB isolates and a correlation with biological factors. *PLoS One* 2018;**13**:e0193237.

Arsenijevic VA, Pekmezovic MG, Meis JF *et al.* Molecular epidemiology and antifungal susceptibility of Serbian *Cryptococcus neoformans* isolates. *Mycoses* 2014;**57**:380–7.

Barchiesi F, Cogliati M, Esposto MC *et al.* Comparative analysis of pathogenicity of *Cryptococcus neoformans* serotypes A, D and AD in murine cryptococcosis. *J Infect* 2005;**51**:10–6.

Bhattacharjee AK, Bennett JE, Glaudemans CPJ. Capsular polysaccharides of *Cryptococcus neoformans*. *Rev Infect Dis* **6**:619–24.

Boekhout T, Theelen B, Diaz M *et al.* Hybrid genotypes in the pathogenic yeast *Cryptococcus neoformans*. *Microbiology* 2001;**147**:891–907.

Casadevall A, Freij JB, Hann-Soden C *et al.* Continental drift and speciation of the *Cryptococcus neoformans* and *Cryptococcus gattii* species complexes. *mSphere* 2017;**2**:e00103-17.

Charlier C, Nielsen K, Daou S *et al.* Evidence of a role for monocytes in dissemination and brain invasion by *Cryptococcus neoformans*. *Infect Immun* 2009;**77**:120–7.

Chowdhary A, Randhawa HS, Sundar G *et al.* In vitro antifungal susceptibility profiles and genotypes of 308 clinical and environmental isolates of *Cryptococcus neoformans* var. *grubii* and *Cryptococcus gattii* serotype B from north-western India. *J Med Microbiol* 2011;**60**:961–7.

Cogliati M. Global Molecular Epidemiology of *Cryptococcus neoformans* and *Cryptococcus gattii*: an atlas of the molecular types. *Scientifica (Cairo)* 2013;**2013**, DOI: 10.1155/2013/675213.

Cogliati M, D'Amicis R, Zani A *et al.* Environmental distribution of *Cryptococcus neoformans* and C. *gattii* around the Mediterranean basin. *FEMS Yeast Res* 2016a;**16**:1–12.

Cogliati M, Prigitano A, Esposto MC *et al.* Epidemiological trends of cryptococcosis in Italy: Molecular typing and susceptibility pattern of *Cryptococcus neoformans* isolates collected during a 20-year period. *Med Mycol* 2018;**56**:963–71.

Cogliati M, Zani A, Rickerts V *et al.* Multilocus sequence typing analysis reveals that *Cryptococcus neoformans* var. *neoformans* is a recombinant population. *Fungal Genet Biol* 2016b;**87**:22–9.

Danesi P, Firacative C, Cogliati M *et al.* Multilocus sequence typing (MLST) and M13 PCR fingerprinting revealed heterogeneity amongst *Cryptococcus* species obtained from Italian veterinary isolates. *FEMS Yeast Res* 2014;**14**:897–909.

Desnos-Ollivier M, Patel S, Raoux-Barbot D *et al.* Cryptococcosis serotypes impact outcome and provide evidence of *Cryptococcus neoformans* speciation. *mBio* 2015;**6**:e00311-15.

Dromer F, Mathoulin S, Dupont B *et al.* Individual and environmental factors associated with infection due to *Cryptococcus neoformans* serotype D. *Clin Infect Dis* 1996;**23**:91–6.

Dromer F, Ronin O, Dupont B. Isolation of *Cryptococcus neoformans* var. *gattii* from an asian patient in France: evidence for dormant infection in healthy subjects. *Med Mycol* 1992;**30**:395–7.

Feng X, Yao Z, Ren D *et al.* Genotype and mating type analysis of *Cryptococcus neoformans* and *Cryptococcus gattii* isolates from China that mainly originated from non-HIV-infected patients. *FEMS Yeast Res* 2008;**8**:930–8.

Freire AKL, dos Santos Bentes A, de Lima Sampaio I *et al.* Molecular characterisation of the causative agents of Cryptococcosis in patients of a tertiary healthcare facility in the state of Amazonas-Brazil. *Mycoses* 2012;**55**:e145–50.

Gago S, Serrano C, Alastruey-Izquierdo A *et al.* Molecular identification, antifungal resistance and virulence of *Cryptococcus neoformans* and *Cryptococcus* de*neoformans* isolated in Seville, Spain. *Mycoses* 2017;**60**:40–50.

Garcia-Hermoso D, Janbon G, Dromer F. Epidemiological evidence for dormant *Cryptococcus neoformans* infection. *J Clin Microbiol* 1999;**37**:3204–9.

General Authority for Statistics Kingdom of Saudi Arabia. *Umrah Survey*., 2017a.

General Authority for Statistics Kingdom of Saudi Arabia. *Haj Statistics*., 2017b.

Goldman D, Lee SC, Casadevall A. Pathogenesis of pulmonary *Cryptococcus neoformans* infection in the rat. *Infect Immun* 1994;**62**:4755–61.

Hagen F, Hare Jensen R, Meis JF *et al.* Molecular epidemiology and in vitro antifungal susceptibility testing of 108 clinical *Cryptococcus neoformans* sensu lato and *Cryptococcus gattii* sensu lato isolates from Denmark. *Mycoses* 2016;**59**:576–84.

Hagen F, Khayhan K, Theelen B *et al.* Recognition of seven species in the *Cryptococcus gattii*/*Cryptococcus neoformans* species complex. *Fungal Genet Biol* 2015;**78**:16–48.

Hagen F, Lumbsch HT, Arsic Arsenijevic V *et al.* Importance of resolving fungal nomenclature: the case of multiple pathogenic species in the *Cryptococcus* genus. *mSphere* 2017;**2**:e00238-17.

Hegazy AK, El-Demerdash MA, Hosni HA. Vegetation, species diversity and floristic relations along an altitudinal gradient in south-west Saudi Arabia. *J Arid Environ* 1998;**38**:3–13.

Husain S, Wagener MM, Singh N. *Cryptococcus neoformans* infection in organ transplant recipients: variables influencing clinical characteristics and outcome. *Emerg Infect Dis* 2001;**7**:375–81.

Kassi FK, Drakulovski P, Bellet V *et al.* Molecular epidemiology reveals genetic diversity among 363 isolates of the *Cryptococcus neoformans* and *Cryptococcus gattii* species complex in 61 Ivorian HIV-positive patients. *Mycoses* 2016;**59**:811–7.

Kavanaugh LA, Fraser JA, Dietrich FS. Recent evolution of the human pathogen *Cryptococcus neoformans* by intervarietal transfer of a 14-gene fragment. *Mol Biol Evol* 2006;**23**:1879–90.

Khayhan K, Hagen F, Pan W *et al.* Geographically structured populations of *Cryptococcus neoformans* variety *grubii* in Asia correlate with HIV status and show a clonal population structure. *PLoS One* 2013;**8**:e72222.

Kinne J, Joseph M, Wernery U *et al.* Disseminated *Cryptococcus deuterogattii* (AFLP6/VGII) infection in an Arabian horse from Dubai, United Arab Emirates. *Rev Iberoam Micol* 2017;**34**:229–32.

Krishna V. Long term temperature trends in four different climatic zones of Saudi Arabia. *Int J Appl Sci Technol* 2014;**4**:233–42.

Kwon-Chung KJ, Bennett JE, Wickes BL *et al.* The case for adopting the "Species Complex" nomenclature for the etiologic agents of Cryptococcosis. *mSphere* 2017;**2**:e00357-16.

Litvintseva AP, Kestenbaum L, Vilgalys R *et al.* Comparative analysis of environmental and clinical populations of *Cryptococcus neoformans*. *J Clin Microbiol* 2005;**43**:556–64.

Litvintseva AP, Mitchell TG. Population genetic analyses reveal the African origin and strain variation of *Cryptococcus neoformans* var. *grubii*. Heitman J (ed.). *PLoS Pathog* 2012;**8**:e1002495.

Mansoor S, Juhardeen H, Alnajjar A *et al.* Hyponatremia as the initial presentation of cryptococcal meningitis after liver transplantation. *Hepat Mon* 2015;**15**:e29902.

Martinez LR, Garcia-Rivera J, Casadevall a. *Cryptococcus neoformans* var. *neoformans* (serotype D) strains are more susceptible to heat than C. *neoformans* var. *grubii* (serotype A) strains. *J Clin Microbiol* 2001;**39**:3365–7.

Meyer W, Aanensen DM, Boekhout T *et al.* Consensus multi-locus sequence typing scheme for *Cryptococcus neoformans* and *Cryptococcus gattii*. *Med Mycol* 2009;**47**:561–70.

Miglia KJ, Govender NP, Rossouw J *et al.* Analyses of pediatric isolates of *Cryptococcus neoformans* from South Africa. *J Clin Microbiol* 2011;**49**:307–14.

Mlinaric-Missoni E, Hagen F, Chew WHM *et al.* In vitro antifungal susceptibilities and molecular typing of sequentially isolated clinical *Cryptococcus neoformans* strains from Croatia. *J Med Microbiol* 2011;**60**:1487–95.

Perfect JR, Dismukes WE, Dromer F *et al.* Clinical practice guidelines for the management of cryptococcal disease: 2010 update by the Infectious Diseases Society of America. *Clin Infect Dis* 2010;**50**:291–322.

Pini G, Faggi E, Bravetti E. Molecular typing of clinical and environmental *Cryptococcus neoformans* strains isolated in Italy. *Open J Med Microbiol* 2017;**07**:77–85.

Rajasingham R, Smith RM, Park BJ *et al.* Global burden of disease of HIV-associated cryptococcal meningitis: an updated analysis. *Lancet Infect Dis* 2017;**17**:873–81.

Rex J, Alexander B, Andes D *et al.* Reference method for broth dilution antifungal susceptibility testing of yeasts: approved standard - third edition. *Clin Lab Stand Inst* 2008;**28**, DOI: 10.4319/lo.1981.26.3.0590.

Sallam AS. Evaluation of some soils in Najd Plateau (Central Region, Saudi Arabia). *J Saudi Soc Agric Sci* 2002;**1**:21–40.

Samarasinghe H, Aceituno-Caicedo D, Cogliati M *et al.* Genetic factors and genotype-environment interactions contribute to variation in melanin production in the fungal pathogen *Cryptococcus neoformans*. *Sci Rep* 2018;**8**:9824.

Sanchini A, Smith IM, Sedlacek L *et al.* Molecular typing of clinical *Cryptococcus neoformans* isolates collected in Germany from 2004 to 2010. *Med Microbiol Immunol* 2014;**203**:333–40.

Tomazin R, Matos T, Meis JF *et al.* Molecular characterization and antifungal susceptibility testing of sequentially obtained clinical *Cryptococcus* d*eneoformans* and *Cryptococcus neoformans* isolates from Ljubljana, Slovenia. *Mycopathologia* 2017;**183**:1–10.

United Nations. *International Migration Report 2017*. New York, USA, 2017.

Del Valle L, Pina-Oviedo S. HIV disorders of the brain; pathology and pathogenesis. *Front Biosci* 2006;**11**:718–32.

Vincent P. *Saudi Arabia: An Environmental Overview*. Taylor & Francis, 2008.

Viviani MA, Cogliati M, Esposto MC *et al.* Molecular analysis of 311 *Cryptococcus neoformans* isolates from a 30-month ECMM survey of cryptococcosis in Europe. *FEMS Yeast Res* 2006;**6**:614–9.

Vogan AA, Khankhet J, Samarasinghe H *et al.* Identification of QTLs associated with virulence related traits and drug resistance in *Cryptococcus neoformans*. *G3 (Bethesda)* 2016;**6**:2745–2759.

Van Wyk M, Govender NP, Mitchell TG *et al.* Multilocus sequence typing of serially collected isolates of *Cryptococcus* from HIV-infected patients in South Africa. *J Clin Microbiol* 2014;**52**:1921–31.

Xu J, Vilgalys R, Mitchell TG. Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*. *Mol Ecol* 2000;**9**:1471–81.

Yan Z, Li X, Xu J. Geographic distribution of mating type alleles of *Cryptococcus neoformans* in four areas of the United States. *J Clin Microbiol* 2002;**40**:965–72.

Chapter 4

Dissection of Melanin QTLs by Bulk Segregant Analysis in *Cryptococcus deneoformans*

4.1 Preface

The ability to produce dark eumelanin pigments is an essential virulence factor of *C. deneoformans* as well as a trait instrumental for their survival in their natural habitats. Significant intraspecific variation in melanin pigmentation is observed among both clinical and environmental *C. deneoformans* isolates. Very few studies have attempted to determine genetic loci underlying this phenotypic variance, with most preferring to focus on its sister species *C. neoformans* instead. Our study is the second quantitative trait loci (QTL) analysis on melanin variance in *C. deneoformans* and the first to apply bulk segregant analysis (BSA) in this species. We identified over 2000 genome-wide single nucleotide polymorphisms as potential melanin QTLs and confirmed differential expression of six candidate genes via quantitative polymerase chain reaction in two strains with divergent melanin phenotypes. Our results highlight the complexity of genetic architecture underlying melanin variance in *C. deneoformans*. We illustrate how BSA can be successfully applied in this species to dissect QTLs of clinically significant phenotypes.

This manuscript is drafted to be published in MDPI Microorganisms. I am the primary contributor of this work. I conducted the majority of the experiments, as well as the analyses and drafting of the manuscript. Co-authors of this study are Yi Lu, Ian Dworkin and Jianping Xu.

4.2 Abstract

The basidiomycetous yeast *Cryptococcus deneoformans* is a causative agent of cryptococcosis with 181,000 deaths worldwide every year. Despite its virulence in humans, *C. deneoformans* is an opportunistic pathogen with its primary reservoir being soil, tree barks and pigeon excreta in the natural environment. Natural *C. deneoformans* strains show significant intraspecific variation in melanin pigmentation, a trait equally essential for virulence and gaining protection from UV light and predatory amoebas in the environment. Here, we applied bulk-segregant analysis (BSA) for the first time in *C. deneoformans* to identify quantitative trait loci (QTL) underlying melanin variance in a lab-derived F1 population. We identified 5 blocks ranging in size from 6kb to 527kb on chromosomes 1, 4 and 7, as well as several, independent single nucleotide polymorphisms (SNP) on chromosomes 3, 5, 6 and 8 as potential melanin QTLs. Six genes located within the QTLs showed differential expression in the two parents, namely, transcription activator GCN5 (CNA03280), ER organization and biogenesis-related protein (CNA04110), low affinity zinc ion transporter (CND00350), calcium ion binding protein (CNG02120), mitotic spindle assembly-related protein (CNG02140), and ligand-regulated transcription factor (CNG02270). Interestingly, neither BSA nor an allele-swap experiment identified *LAC1*, which encodes the laccase enzyme essential for melanin production in *C. deneoformans,* as a potential melanin QTL. Our findings suggest that different genes and regulatory elements might play crucial roles in melanin production vs. quantitative variance in pigmentation. Our study highlights the complex genetic architecture underlying phenotypic variance in the medically relevant yeast *C. deneoformans.* Its implications for virulence and ongoing evolution amidst a changing climate remain to be investigated.

4.3 Introduction

Cryptococcosis is an AIDS-defining illness that affects over 220 000 individuals worldwide every year, with 181 000 deaths [1]. Cryptococcosis can take a variety of forms including skin infections, lung infections and meningoencephalitis [2–4]. The causative agent of this deadly infection is a pair of closely related, basidiomycetous yeast species, *Cryptococcus neoformans* and *Cryptococcus deneoformans*. While *C.*

neoformans is considered the more virulent species with a global distribution, *C. deneoformans* is responsible for up to 40% of cryptococcal infections in Mediterranean Europe [5–7]. A recent multi-species animal specimen study from Poland indicated that *C. deneoformans* might be the dominant species (74%) in animal reservoirs including feral pigeons, parrots and lemurs [8]. Ability to synthesize black/brown eumelanin pigments from exogenous substrates is one of three 'dual-role' phenotypes (the other two being the production of a polysaccharide cell capsule and the ability to grow at 37°C) that confers these yeasts resistance to common stressors in their natural environment such as UV radiation, heat, and predation, as well as protection from the host's immune system during infections [9–11].

Significant intraspecific variation in melanin pigmentation is observed among environmental and clinical strains of *C. deneoformans* [12]. Given the survival benefits conferred by melanin pigments, reasons for such variance remain unclear. There could be fitness trade-offs between vegetative growth and the metabolically expensive process of melanin biosynthesis [13]. Environmental strains in different geographical locations could experience different microclimates and differential exposure to environmental stressors, requiring different amounts of melanisation. Previous work from my group identified genotype-environment interactions as a significant contributor to melanin variance in a global *C. deneoformans* population, suggesting that melanin biosynthesis is responsive and adaptable to the level of stress in the environment [12].

Melanin biosynthesis is a complex metabolic process in *C. deneoformans,* requiring the functionality of a multitude of enzymes and a network of regulatory elements [13]. A phenoloxidase, encoded by the *LAC1* gene located on chromosome 7, is essential for catalyzing the initial oxidation step of the substrate, typically L-3,4-dihydroxyphenylalanine (L-DOPA), into dopaquinone, which kicks off a series of spontaneous reactions that ends in the polymerization of melanin pigments [14]. Two major signaling pathways have been implicated in the regulation of *LAC1* expression and melanin production in *C. deneoformans,* namely the cyclic AMP/protein kinase A (cAMP/PKA) and high-osmolarity glycerol response (HOG) pathways [15–18]. Transcription factors and kinases associated with these pathways, such as Bzp4,

Usv101, Mbs1, Hob1, GSK3 and KIC1, have been implicated in the regulation and fine-tuning of *LAC1* expression [19]. While gene deletion experiments established the critical roles these components play in melanin biosynthesis, it is unclear if and how they contribute to quantitative variance in melanisation observed within *C. deneoformans* populations.

We previously identified nucleotide polymorphisms in the *LAC1* gene to be significantly correlated with the level of melanin pigmentation in *C. deneoformans* [12]. A recent quantitative trait locus (QTL) mapping using whole genome sequences of progeny derived from crosses between *C. deneoformans* strains identified *RIC8,* which encodes a regulator in the cAMP/PKA signaling pathway, as contributing to melanin variance [20]. A previous QTL analysis using polymerase chain reaction-restriction fragment length polymorphisms (PCR-RFLP) markers in a hybrid progeny population derived from a cross between *C. neoformans* and *C. deneoformans* identified five, putative QTLs contributing to melanin variance, which included an arginine n-methyltransferase located on chromosome 2, a stomatin-like protein on chromosome 14, and *LAC1* [21]. In 2006, Lin and colleagues identified *MAC1,* a putative ligand-regulated transcription factor located on chromosome 7, as a candidate QTL contributing to melanin variance [22]. The percentage of melanin variance explained by the QTLs identified in the above studies in their respective populations ranged from 10% to 67.5%, indicating that the full genetic architecture behind this intraspecific variance remains to be elucidated.

Bulk segregant analysis (BSA) is a form of QTL mapping where DNA from individuals displaying transgressive phenotypes is pooled and subjected to whole genome sequencing, with alleles enriched in one pool over the other identified as putative QTLs contributing to phenotypic variance. In closely related, progeny populations, pooling DNA of individuals with similar phenotypes can enrich alleles contributing to extreme phenotypes, making BSA an attractive technique to be used in the dissection of QTLs in F1 populations. BSA has been successfully implemented in multiple studies on the model yeast *Saccharomyces cerevisiae* to identify QTLs contributing to variance in traits of clinical and industrial importance including cantharidin resistance, virulence in *Galleria mellonella,* high ethanol tolerance, xylose

utilisation, thermotolerance and lignocellulosic inhibitor tolerance, among others [23–28]. In the model fungus *Neurospora crassa*, researchers have used BSA to identify genes involved in cell cycle regulation and programmed cell death [29,30].

Here we describe the application of BSA for the first time in *C. deneoformans* to identify melanin QTLs in a F1 population of 425 strains derived in the lab from a cross between two strains with divergent melanin phenotypes. We assess gene expression levels of top candidate genes in the parental strains using quantitative polymerase chain reaction (qPCR). We also verify the contribution of three single nucleotide polymorphisms (SNPs) in the *LAC1* gene to melanin variance by conducting an allele swap between two strains with differential melanin production.

4.4 Results

Melanin variance in F1 population

We obtained 425 F1 progeny strains from our laboratory cross between JEC21 and NIH-433. We observed significant variance in melanin production among strains in the F1 population (Figure 1). Quantitative data from the spot densitometry assay showed the presence of transgressive melanin phenotypes in the progeny population: 23 progeny strains produced more melanin than the darker parent NIH-433, while 209 strains produced less melanin than the lighter parent JEC21, 54 of which did not produce any pigments at all (i.e. albino strains). The remaining 139 progeny strains produced an intermediate amount of melanin between the two parents. Our results confirmed that melanin pigmentation was a quantitative trait in this *C. deneoformans* F1 population.

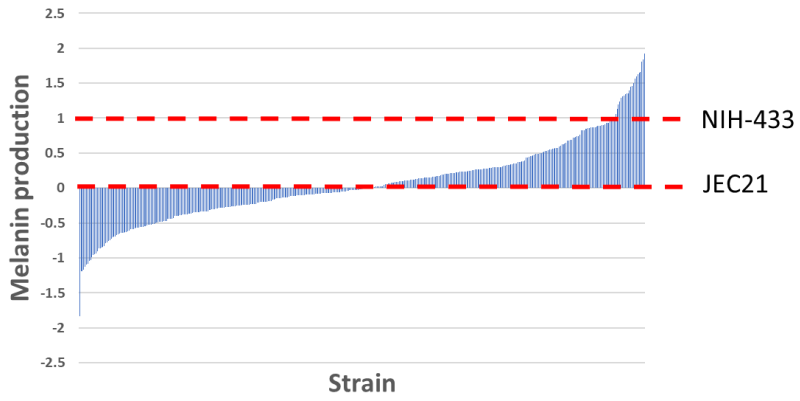


Figure 1: Melanin production of the F1 population. The 425 progeny strains obtained from JEC21 x NIH-433 cross showed a wide range in melanin production. Melanin production of progeny strains is shown here in reference to that of the parents: melanin levels of JEC21 and NIH-433 were adjusted to zero and one respectively.

Bulk Segregant Analysis

We selected the 25 highest and 25 lowest melanin producing progeny strains to create the high and low melanin pools, respectively. We performed whole genome sequencing on the two pools and the parental strains at approximately 100x coverage. Following filtering, we identified 50,706 bi-allelic SNPs between the two parents. We determined the genotypes of the two pools at these sites. The SNPs were not uniformly distributed across the genome: most chromosomes had non-variable blocks interspersed between regions of high SNP density (Figure 2). The two parental strains used in our cross are related to each other. NIH-433 is a progenitor strain used in the construction of the laboratory strain, JEC21, which explains the shared blocks of chromosomes in their genomes [31].

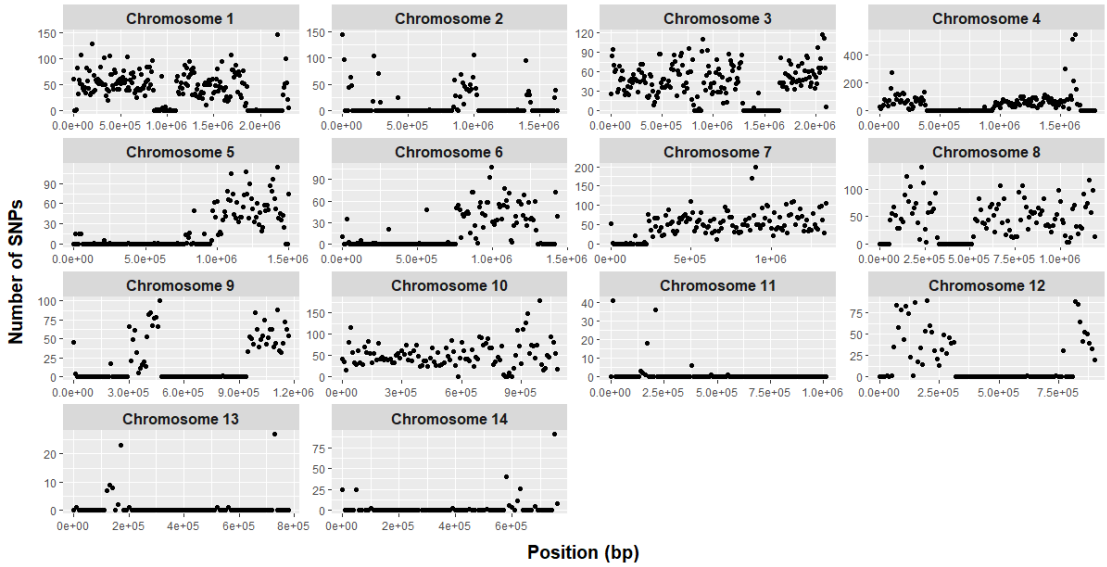


Figure 2: Genome-wide SNP density in 10kb windows. In total, 50706 high-quality SNPs were identified between the two parental strains, JEC21 and NIH-433. SNPs were not uniformly distributed across the genome.

We calculated the F_{ST} of each SNP as a measure of differentiation between the two pools. F_{ST} values can range from 0 to 1 where 1 indicates complete differentiation (i.e. one pool exclusively contains one parental allele while the second pool has the other parent's allele) and 0 indicates no differentiation (both pools have both parental alleles at similar frequencies). F_{ST} values ranging the full spectrum from 0 to 1 were observed in our SNP dataset, with 2196 SNPs having a F_{ST} value greater than 0.4, which indicates a moderate level of differentiation (Figure 3). High F_{ST} SNPs were predominantly found in blocks among other high F_{ST} SNPs. Chromosome 1 contained a 62kb block and a 294kb block spanning 845,630bp - 907,259bp and 1,032,506bp – 1,326,185bp respectively. Chromosome 4 contained two blocks, with the first one spanning 6kb from 1,293,428bp to 1,299,376bp, and the second spanning 86kb from 1,548,735bp to 1,634,829bp. A 527kb block was found on chromosome 7 spanning from 438,478bp to 965,397bp. The remaining high F_{ST} SNPs were found scattered across chromosome 3 (five SNPs), chromosome 5 (two SNPs), chromosome 6 (two SNPs), and chromosome 8 (23 SNPs).

We annotated our full SNP dataset using the published annotations of the JEC21 reference genome on NCBI [32]. The annotation summary of the 2196 SNPs with F_{ST} greater than 0.4 is listed in Table 1. We considered SNPs located within 1500bp

upstream of a gene to be upstream variants with potential to affect gene expression. The 2196 SNPs were located in 218 genes with predicted effects ranging from synonymous amino acid changes to alternations in gene expression and protein structure. We established the putative identity and functionality of 133 of the 218 genes based on published annotations, listed in Supplementary Table 1. 29 genes were located on chromosome 1, two on chromosome 3, eight on chromosome 4, 90 on chromosome 7, and four on chromosome 8. Interestingly, the high F_{ST} block on chromosome 7 did not include the *LAC1* gene, which was located ~83kb upstream of the block. The remaining 85 genes were recognized as 'hypothetical proteins' or 'expressed proteins', with no further details available on their function.

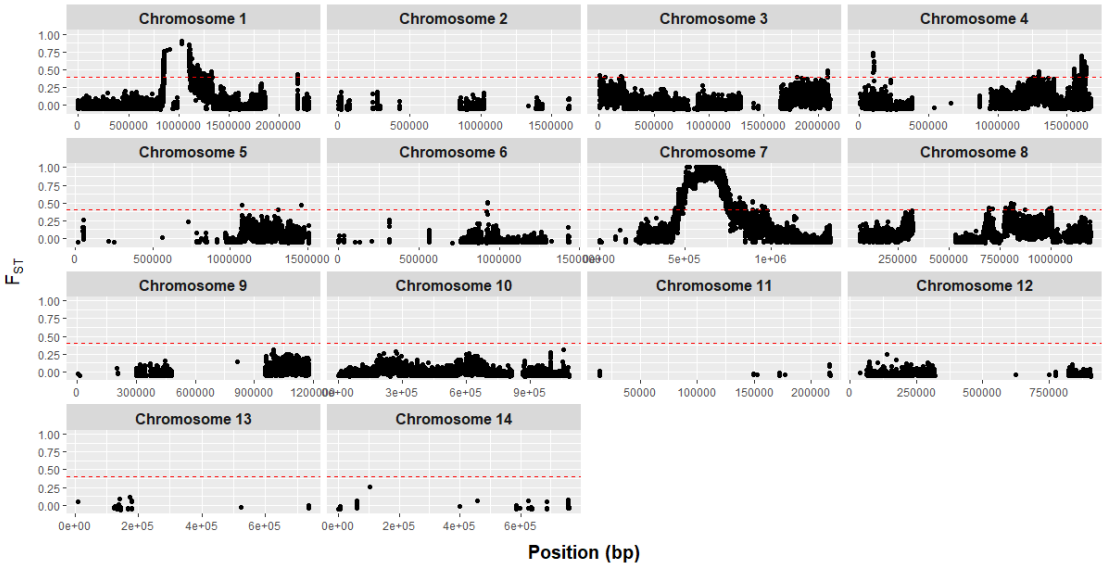


Figure 3: F_{ST} of genome-wide SNPs. In this dataset, F_{ST} values range from 0 to 1. Most SNPs with high F_{ST} values occur in three blocks on chromosomes 1 and 7.

Table 1: Summary of annotations of the 2196 SNPs with a F_{ST} value greater than 0.4.

Predicted effect	Number of SNPs
Synonymous variant	772
Missense variant	369
Intron variant	334
Intergenic region	254
5' UTR variant	154
3' UTR variant	144
Splice region variant and intron variant	78
Intragenic variant	38
5' UTR premature start codon gain variant	16
Non-coding transcript variant	10
Splice region variant and synonymous variant	10
Missense variant and splice region variant	8
Splice region variant	4
Splice acceptor variant and intron variant	1
Splice donor variant and intron variant	1
Start lost	1
Stop lost	1
Stop retained variant	1
Total	**2196**

Gene expression of candidate genes

We used RT-qPCR to quantify the expression of six candidate genes in the two parental strains. We selected CNA03280 (transcriptional activator gcn5) and CNA04110 (ER organization and biogenesis-related protein, putative) located within the 250kb block of high F_{ST} SNPs on chromosome 1, CND00350 (low-affinity zinc ion transporter, putative) associated with a high F_{ST} SNP on chromosome 4, and CNG02120 (calcium ion binding protein, putative), CNG02270 (ligand-regulated transcription factor, putative), and CNG02140 (mitotic spindle assembly -related protein, putative), which were located within the 259kb fragment of high F_{ST} SNPs on chromosome 7. All six genes showed significantly differential expression between the

two parental strains under melanized conditions (Figure 4). The expression of the calcium ion binding protein was significantly higher in NIH-433, the high melanin producing parent, while the expression of the remaining five genes was significantly higher in JEC21, the lower melanin producer.

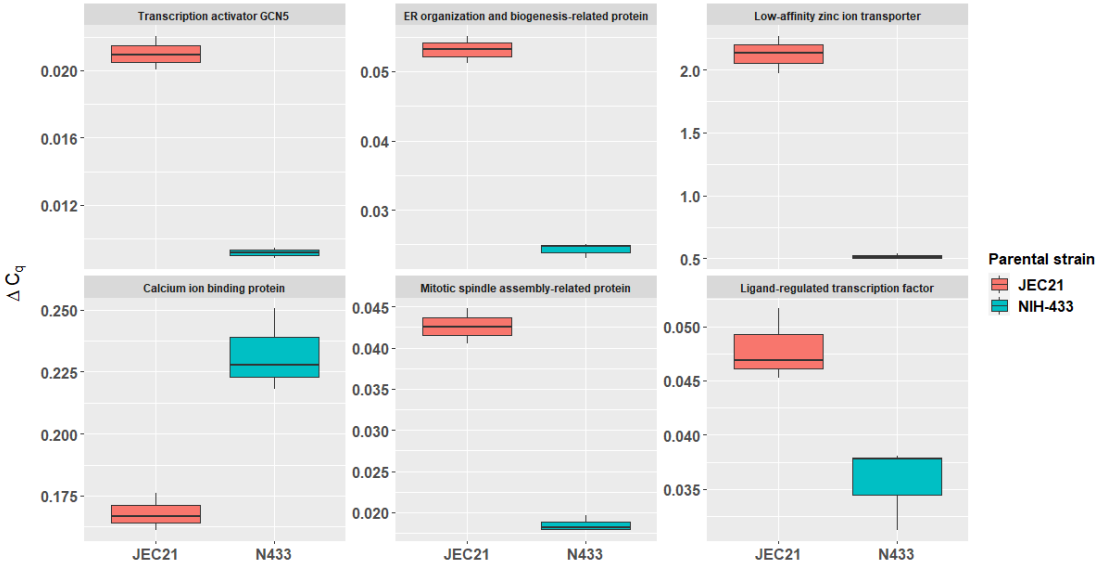


Figure 4 Differential gene expression of six candidate genes in the two parental strains. Gene expression was quantified using RT-qPCR and differential expression was calculated using the delta C_q method.

LAC1 allele swap

To determine the contribution of genetic polymorphisms of the *LAC1* locus to melanin variance, we used CRISPR-Cas9 to conduct an allele swap between JEC21 and GRLMM26HO1–2, an environmental strain from Greece that produces significantly more melanin at 30°C. The two strains differ at the *LAC1* locus by several SNPs: three SNPs are located within introns, while 13 SNPs are found within exons, six of which cause non-synonymous amino acid changes. We replaced a 853-bp fragment within the first half of the *LAC1* coding region in JEC21 with that of GRLMM26HO1–2, which incorporated three non-synonymous SNPs (Figure 5). The resulting recombinant strain, named GJ1, contained a recombinant *LAC1* gene in a JEC21 genetic background. We quantified the melanin production of GJ1, JEC21 and GRLMM26HO1–2 at both 30°C and 37°C. As expected, GRLMM26HO1–2 produced significantly more melanin than JEC21 at both temperatures. GJ1 was

indistinguishable from the recipient strain, JEC21, in melanin production, suggesting that the three SNPs altered in the allele swap do not contribute to melanin variance observed between JEC21 and GRLMM26HO1–2 (Figure 6).

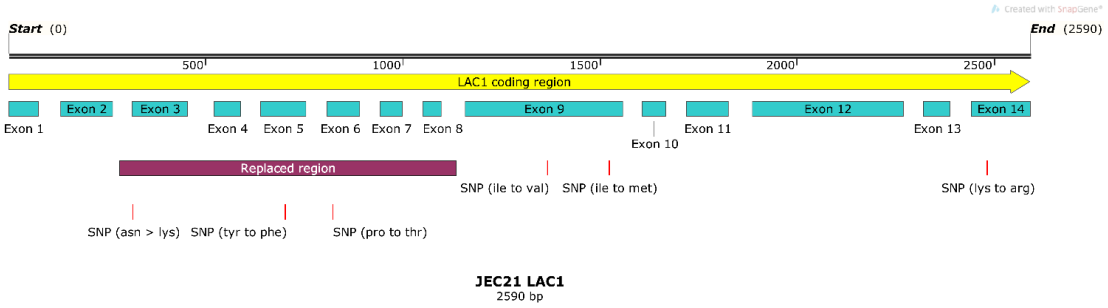


Figure 5: *LAC1* gene codes for a laccase essential for melanin production in *Cryptococcus deneoformans*. *LAC1* coding region is 2590bp long and consists of 14 exons. The six non-synonymous SNPs between recipient strain, JEC21 and donor strain, GRLMM26HO1–2 are marked in red and the amino acid changes are noted within parantheses. We replaced a 853bp region in JEC21 with that of GRLMM26HO1–2.

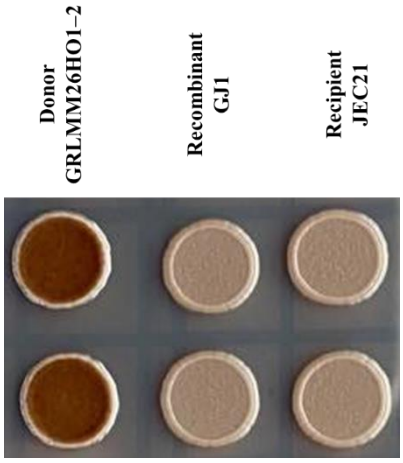


Figure 6 Melanin production of GRLMM26HO1–2, GJ1 and JEC21 at 30°C. The first half of *LAC1* in JEC21 was replaced with that of GRLMM26HO1–2. The resulting recombinant strain, GJ1 with the *NAT* resistance marker, was indistinguishable from JEC21 in melanin pigmentation.

4.5 Discussion

Melanin biosynthesis in *C. neoformans* and *C. deneoformans* is a complex biochemical process, requiring careful coordination between synthesis, transport, and cell wall deposition of melanin pigments (reviewed in [13]). Each step is mediated and regulated by a multitude of regulatory elements and major signaling pathways, making melanin biosynthesis one of the most complex metabolic processes to take place in these yeast cells. Theoretically, any of the hundreds of genes involved in melanization

could contribute to intraspecific variation in pigmentation, requiring broad-scale, genetic techniques to map melanin QTLs. In this study, we applied BSA for the first time in *C. deneoformans* to dissect genetic loci associated with melanin variance. Our results confirm the quantitative nature of melanin production in *C. deneoformans*, with mating between closely related strains generating offspring with significant phenotypic diversity.

Most previous studies investigating genetic, molecular, and biochemical properties of melanisation have focused on *C. neoformans*, with a distinct lack of similar studies conducted on the sister species *C. deneoformans*. Given the 10-15% differentiation in nucleotide identity between the two species, the genetic architecture and dynamics behind phenotypic variance in virulence factors could also differ between them [33–35]. The recent QTL analysis of major virulence factors by Roth and colleagues is the only study we are aware of that elucidated melanin QTLs in *C. deneoformans* specifically [20]. In the current study, we identified 218 genes associated with 2196 genome-wide SNPs as potential melanin QTLs in *C. deneoformans*. *RIC8*, a guanine nucleotide exchange factor for Gpa1, identified by Roth and colleagues as a melanin QTL was not flagged as a candidate QTL in our mapping population. The two studies differed in the parental strains used and how the mapping population was generated. Roth et al. combined segregants from two crosses, one unisexual and one bisexual, to generate their mapping population. Strains 431α, XL280αSS, and XL280a were used as parents, with the latter two being congenic strains, differing only at the mating locus, *URA5* gene and a partial duplication of the left arm of chromosome 10. Our mapping population was drawn from a single mating event between two, related but non-congenic strains. Findings of the two QTL studies suggest that gene expression dynamics underlying phenotypic variance can be highly dependent on the genetic background.

Our BSA analysis revealed 5 blocks on chromosomes 1, 4 and 7, ranging in size from 6kb to 527kb, that were significantly differentiated between high and low melanin pools. Together, these blocks contained 207 genes, with putative functionality identified for 127, each of which is a potential melanin QTL in our population. Since the block on chromosome 7 did not include *LAC1*, genetic polymorphisms at *LAC1*

are likely not a significant contributor to melanin variance in this F1 population. The appearance of large blocks of high F_{ST} SNPs could be due to lack of recombination in these regions between the parental chromosomes. A conservative estimate of the number of QTLs identified in this study would be 19, accounting for one gene per each of the five blocks, and three, two, one and eight genes associated with individual, high F_{ST} SNPs located on chromosomes 3, 5, 6 and 8 respectively. Since the two parental strains share ~50% genetic ancestry, the true number of QTLs could be double the estimated number here.

A previous study identified 27 transcription factors (TF) in the *C. neoformans* clinical strain H99, which, when deleted, led to an increase or decrease in melanin production [36]. We identified homologs of these TFs in the JEC21 genome via BLAST searches: homologs of three TFs, namely *SKN7*, *CUF1* and *BZP4*, were located within the block on chromosome 7, suggesting that these TFs could be contributing to melanin variance in *C. deneoformans*. A later study identified Bzp4 as one of four core TFs involved in the regulation of melanin production in *C. neoformans.* Our results indicate that Bzp4 may play a critical role in not only the regulation but variance of melanin pigmentation between *C. deneoformans* strains as well. A previous functional analysis of kinases in *C. neoformans* strain H99 identified 26 kinases, when deleted, led to melanin production being enhanced or reduced [37]. When we identified homologs of these genes in the JEC21 genome via BLAST searches, we determined that genes of three kinases, namely *PKA1*, *KIC1* and *CDC2801*, were located within the 294kb block on chromosome 1 while another kinase, *TCO6*, was located within the 62kb block on the same chromosome. Overall, our results suggest that homologs of three TFs and four kinases, previously identified in *C. neoformans* as involved in regulation of melanin synthesis, contribute to intraspecific melanin variance in *C. deneoformans.*

Our RT-qPCR assays confirmed differential expression of six candidate genes in melanized parental strains, implicating them as melanin QTLs contributing to phenotypic variance in our *C. deneoformans* mapping population. Two are regulators of gene expression, namely histone acetyltransferase *GCN5* (CNA03280) and a ligand-regulated transcription factor (CNG02270). While melanin production was not

affected when *GCN5* was disrupted in the *C. neoformans* strain H99, other studies found melanin production to be affected when *C. neoformans* strains were grown in the presence of drugs that inhibited histone deacetylases like GCN5 [38,39]. The third novel melanin QTL we identified was a low-affinity zinc ion transporter (CND00350) located on chromosome 4. Zinc ion transporters were shown to be essential for virulence in the closely related, pathogenic yeast, *Cryptococcus gattii*, whose melanin production was significantly reduced when the zinc ion transporter, Zip3, was disrupted [40]. The fourth QTL was annotated as a calcium ion binding protein (CNG02120) located on chromosome 7. *C. neoformans* cells rely on calcium ions for correct vacuole acidification which is essential for its virulence, with impaired vacuolar activity resulting in loss of virulence, as well as decreased production of melanin production in *C. neoformans* strain H99 [41,42]. The ER organization and biogenesis-related protein (CNA04110) and mitotic spindle assembly-related protein (CNG02140), located on chromosomes 1 and 7 respectively, were also identified by our gene expression assays as putative, novel melanin QTLs. The contribution of these genetic loci to melanin variance in *C. deneoformans* should be further validated via gene knockout and complementation experiments.

BSA is an experimental approach commonly applied in the model yeast *S. cerevisiae* to study the genetic architecture of complex, quantitative traits. BSA, when combined with functional validation assays, can be a powerful tool to identify hundreds of QTLs contributing to a phenotype in a short amount of time. A recent study identified 51 and 96 QTLs affecting growth and survival respectively in *S. cerevisiae*, with downstream experiments confirming the identity of three, novel genes contributing to the phenotypes [43]. Our combined BSA and gene expression assays identified six, putative melanin QTLs that have not been implicated in melanin variance in any previous studies. Our study highlights the complex genetic backdrop of melanin production, an essential virulence factor, in *C. deneoformans*, and opens up the opportunity to expand our understanding of cryptococcal biology and virulence.

4.6 Materials and Methods

Strains, media, and generation of F1 progeny

We selected JEC21 (mating type α), a low melanin producing, standard laboratory strain, and NIH-433 (mating type a), a high melanin producing environmental strain isolated from a pigeon nest in Denmark, as parental strains. The two strains share ~50% genetic ancestry, due to JEC21 being congenic to JEC20 who was the meiotic progeny of a cross between NIH-433 and NIH-12 [44]. In addition to their phenotypic variance in melanin production, JEC21 and NIH-433 differ at the *LAC1* locus by several single nucleotide polymorphisms that were previously shown to be significantly associated with melanin variance [45]. We mixed equal amounts of the two strains in cell suspensions of 1 x 10^5 cells/ml in nuclease-free water. We spotted 10ul of the cell mixture on to Murashige and Skoog (MS) agar medium with an adjusted pH of 5. We prepared several mating plates with multiple mating spots and incubated them at 23°C in the dark for ~4 weeks until robust hyphal growth was visible around the colonies. To collect basidiospores, we cut out a small square of agar containing hyphae using a sterile scalpel, being careful not to touch the colony with parental yeast cells. The agar piece was crushed using a sterile, blue stick in a 2ml Eppendorf tube, suspended in 1ml of 0.2% Tween solution and vortexed for 5 minutes with 200ul of 0.6mm glass beads to extract spores from agar into the solution. This solution was passed through a syringe filter with a pore size of 5µm pore size (Millipore Sigma catalog no: SLSV025LS) to further exclude any parental yeast and hyphal cells. The filtered solution was spread onto Yeast Extract-Potato-Dextrose (YEPD) agar plates and incubated for 2-3 days at 30°C until germinated colonies appeared. Colonies were sub-cultured onto fresh YEPD plates and streaked for single colonies to obtain pure F1 progeny strains.

Melanin quantification assay

We quantified the melanin production of parental and progeny strains on caffeic acid plates at 30°C using spot densitometry, following the protocol outlined in [45].

Bulk segregant analysis

We combined the 25 highest and 25 lowest melanin producing progeny strains in our F1 population to form the high and low melanin pools respectively. We grew the selected 50 strains on YEPD for 3 days to obtain fresh cells. We suspended cells in nuclease-free water and used a cell counter to adjust the concentration to 1 x 10^6 cells/ml. For each pool, we combined equal volumes of the 25 cell suspensions to obtain a cell mixture with each strain equally represented. We extracted genomic DNA from the two cell mixtures (i.e. high and low melanin pools) as well as from the two parental strains using the standard chloroform-isoamyl alcohol method, outlined in [21]. We performed Illumina sequencing on a Hiseq 2000 platform at 90x coverage for the parental strains and 100x coverage for the pools. Sequencing was performed at The Center for Applied Genomics in Toronto, Canada.

Bioinformatics workflow

After performing quality checks on raw read files using FastQC and MultiQC [46], we trimmed adapter sequences off the raw reads using Trimmomatic [47]. We mapped the raw reads of our four samples (JEC21, NIH-433, high melanin pool and low melanin pool) to the JEC21 reference genome available on NCBI (Genbank assembly accession: GCA_000091045.1) using the Burrows-Wheeler Aligner [48]. We sorted and removed duplicate reads from the resulting BAM files using Picard Tools. We used the VarScan variant caller to identify genetic polymorphisms (SNPS and Insertion/Deletions) in our four samples against the JEC21 reference genome [49]. We used GATK tools to extract SNPs from the variants file which were then hard filtered to remove low-quality SNPs [50]. Next, we used SnpEff [51] to annotate the SNPs based on the published JEC21 gene annotations. We used the poolfstat package [52] in R to calculate the f-statistic (Fst) for each SNP which is a measure of its differentiation between the two pools.

RT-qPCR to determine gene expression of potential QTLs

First, we extracted SNPs from the full dataset that were predicted to be located upstream of genes and could affect their expression. Within this subset, we selected the SNPs with the highest Fst values and identified the genes located immediately downstream of them. We selected six genes with reliable annotations of identity and

function as potential QTLs: we avoided picking genes that were simply annotated as hypothetical proteins. We used the IDT PrimerQuest™ Tool to design qPCR primers spanning exon-exon junctions of the mRNA of the six genes. We performed a melanin assay on the two parental strains and extracted total RNA after 7 days' incubation following the protocol outlined in [12]. We performed reverse transcriptase-qPCR (RT-qPCR) on extracted RNA samples using the Luna University One-Step RT-qPCR kit (NEB catalog no: E3005X) to quantify the mRNA levels of the six, candidate genes in the parental strains when melanized. In all RT-qPCR assays, we used actin as the reference gene. Primers used for RT-qPCR are listed in Supplementary Table 3.

LAC1 allele swap using CRISPR-Cas9

In a previous study, we identified SNPs located within the *LAC1* coding region that were significantly correlated with melanin variance among 54 natural *C. deneoformans* strains [12]. We selected GRLMM26HO1–2, an environmental strain isolated from an Olive tree in Greece, as the allele donor in the *LAC1* allele swap experiment due to its high melanin production at 30°C. We replaced the first half of *LAC1'*s coding region in JEC21 with that of GRLMM26HO1–2, which included three non-synonymous SNPs, using Transient CRISPR-Cas9 Coupled with Electroporation (TRACE) technique, as outlined in [53,54]. In the same TRACE reaction, we introduced two DNA constructs into the recipient cells: i) *LAC1* allele swap construct, ii) *NEO* construct.

We built the *LAC1* construct by using NEBuilder® HiFi DNA Assembly Master Mix (NEB catalog no: E2621X) to ligate the *LAC1* allele fragment from the donor strain and the 5' and 3' arms from the recipient strain into the linearized plasmid vector, pBSKS-. The assembled product was transformed into NEB 10-beta competent *E. coli* cells (NEB catalog no: C3019) following manufacturer's transformation protocols. We introduced the geneticin (G418) resistance marker, *NEO,* using linearized plasmid pCF3, to the Safe Haven genomic region during the same TRACE reaction to facilitate selection of successful transformants [55]. We used Q5 Hot Start High-Fidelity 2X Master Mix (NEB catalog no: M0494X) for all polymerase chain reactions (PCR). Primers we designed for our TRACE reactions are listed in

Supplementary Table 3. Electroporation was carried out using a BioRad MicroPulser Electroporator at 2000V (time constant was not adjustable).

We selected one transformant that was resistant to G418 and confirmed the successful allele swap by sequencing the full *LAC1* gene. This recombinant strain was named GJ1. We quantified the melanin production of GJ1, along with donor and recipient strains, using our melanin quantification assay, mentioned above.

4.7 Supplementary Data

Supplementary Table 1: Annotations and F_{ST} values of 2196 SNPs with F_{ST} greater than 0.4

Please click on the following link to access the table.

https://docs.google.com/spreadsheets/d/1pNLPslylzB2pvw01lgSTSYFPALGXFDSm/edit?usp=sharing&ouid=106686285052112798692&rtpof=true&sd=true

Supplementary Table 2: Annotations of 218 genes associated with the high F_{ST} SNPs

Please click on the following link to access the table.

https://docs.google.com/spreadsheets/d/1kElXx_egBSFKhKN6p3vPR0eMsMNyMGkZ/edit?usp=sharing&ouid=106686285052112798692&rtpof=true&sd=true

Supplementary Table 3: Primers used in RT-qPCR and TRACE

Please click on the following link to access the table.

https://docs.google.com/spreadsheets/d/1_YtOUb2poVsRxF6n-QSRZYYti6xffoP_/edit?usp=sharing&ouid=106686285052112798692&rtpof=true&sd=true

4.8 References

1. Rajasingham, R.; Smith, R.M.; Park, B.J.; Jarvis, J.N.; Govender, N.P.; Chiller, T.M.; Denning, D.W.; Loyse, A.; Boulware, D.R. Global burden of disease of HIV-associated cryptococcal meningitis: an updated analysis. *Lancet Infect. Dis.* **2017**, *17*, 873–881.

2. Goldman, D.; Lee, S.C.; Casadevall, A. Pathogenesis of pulmonary Cryptococcus neoformans infection in the rat. *Infect. Immun.* **1994**, *62*, 4755–4761.

3. Husain, S.; Wagener, M.M.; Singh, N. Cryptococcus neoformans infection in organ transplant recipients: variables influencing clinical characteristics and outcome. *Emerg. Infect. Dis.* 2001, *7*, 375–381.

4. Charlier, C.; Nielsen, K.; Daou, S.; Brigitte, M.; Chretien, F.; Dromer, F. Evidence of a role for monocytes in dissemination and brain invasion by Cryptococcus neoformans. *Infect. Immun.* **2009**, *77*, 120–127.

5. Cogliati, M. Global Molecular Epidemiology of Cryptococcus neoformans and Cryptococcus gattii: An Atlas of the Molecular Types. *Scientifica (Cairo).* **2013**, *2013*.

6. Bovers, M.; Hagen, F.; Boekhout, T. Diversity of the Cryptococcus neoformans-Cryptococcus gattii species complex. *S4 Rev Iberoam Micol* **2008**, *25*, 4–12.

7. Cogliati, M.; Desnos-Ollivier, M.; McCormick-Smith, I.; Rickerts, V.; Ferreira-Paim, K.; Meyer, W.; Boekhout, T.; Hagen, F.; Theelen, B.; Inácio, J.; et al. Genotypes and population genetics of cryptococcus neoformans and cryptococcus gattii species complexes in Europe and the mediterranean area. *Fungal Genet. Biol.* **2019**, *129*, 16–29.

8. Florek, M.; Nawrot, U.; Korzeniowska-Kowal, A.; Włodarczyk, K.; Wzorek, A.; Woźniak-Biel, A.; Brzozowska, M.; Galli, J.; Bogucka, A.; Król, J. An analysis of the population of Cryptococcus neoformans strains isolated from animals in Poland, in the years 2015–2019. *Sci. Reports 2021 111* **2021**, *11*, 1–

12.

9. Casadevall, A.; Steenbergen, J.N.; Nosanchuk, J.D. “Ready made” virulence and “dual use” virulence factors in pathogenic environmental fungi - the Cryptococcus neoformans paradigm. *Curr. Opin. Microbiol.* **2003**, *6*, 332–337.

10. Casadevall, A.; Rosas, A.L.; Nosanchuk, J.D. Melanin and virulence in Cryptococcus neoformans. *Curr. Opin. Microbiol.* **2000**, *3*, 354–358.

11. Schultzhaus, Z.; Chen, A.; Kim, S.; Shuryak, I.; Chang, M.; Wang, Z. Transcriptomic analysis reveals the relationship of melanization to growth and resistance to gamma radiation in Cryptococcus neoformans. *Environ. Microbiol.* **2019**, *21*, 2613–2628.

12. Samarasinghe, H.; Aceituno-Caicedo, D.; Cogliati, M.; Kwon-Chung, K.J.; Rickerts, V.; Velegraki, A.; Ackaglar, S.; Xu, J. Genetic Factors and Genotype-Environment Interactions Contribute to Variation in Melanin Production in the Fungal Pathogen Cryptococcus neoformans. *Sci. Rep.* **2018**, *8*, 9824.

13. Cordero, R.J.B.; Camacho, E.; Casadevall, A. Melanization in cryptococcus neoformans requires complex regulation. *MBio* **2020**, *11*.

14. Eisenman, H.C.; Casadevall, A. Synthesis and assembly of fungal melanin. *Appl. Microbiol. Biotechnol.* 2012, *93*, 931–940.

15. Alspaugh, J.A.; Perfect, J.R.; Heitman, J. Cryptococcus neoformans mating and virulence are regulated by the G-protein α subunit GPA1 and cAMP. *Genes Dev.* **1997**, *11*, 3206–3217.

16. Bahn, Y.S.; Hicks, J.K.; Giles, S.S.; Cox, G.M.; Heitman, J. Adenylyl cyclase-associated protein Aca1 regulates virulence and differentiation of Cryptococcus neoformans via the cyclic AMP-protein kinase A cascade. *Eukaryot. Cell* **2004**, *3*, 1476–1491.

17. Bahn, Y.-S.; Kojima, K.; Cox, G.M.; Heitman, J. Specialization of the HOG Pathway and Its Impact on Differentiation and Virulence of Cryptococcus neoformans. *https://doi.org/10.1091/mbc.e04-11-0987* **2005**, *16*, 2285–2300.

18. Bahn, Y.S.; Geunes-Boyer, S.; Heitman, J. Ssk2 mitogen-activated protein kinase kinase kinase governs divergent patterns of the stress-activated Hog1 signaling pathway in Cryptococcus neoformans. *Eukaryot. Cell* **2007**, *6*, 2278–2289.

19. Lee, D.; Jang, E.H.; Lee, M.; Kim, S.W.; Lee, Y.; Lee, K.T.; Bahna, Y.S. Unraveling melanin biosynthesis and signaling networks in cryptococcus neoformans. *MBio* **2019**, *10*.

20. Roth, C.; Murray, D.; Scott, A.; Fu, C.; Averette, A.F.; Sun, S.; Heitman, J.; Magwene, P.M. Pleiotropy and epistasis within and between signaling pathways defines the genetic architecture of fungal virulence. *PLOS Genet.* **2021**, *17*, e1009313.

21. Vogan, A.A.; Khankhet, J.; Samarasinghe, H.; Xu, J. Identification of QTLs Associated with Virulence Related Traits and Drug Resistance in Cryptococcus neoformans. *G3 (Bethesda).* **2016**, *6*, 2745–2759.

22. Lin, X.; Huang, J.C.; Mitchell, T.G.; Heitman, J. Virulence attributes and hyphal growth of C. neoformans are quantitative traits and the MATα allele enhances filamentation. *PLoS Genet.* **2006**.

23. Wilkening, S.; Lin, G.; Fritsch, E.S.; Tekkedil, M.M.; Anders, S.; Kuehn, R.; Nguyen, M.; Aiyar, R.S.; Proctor, M.; Sakhanenko, N.A.; et al. An Evaluation of High-Throughput Approaches to QTL Mapping in Saccharomyces cerevisiae. *Genetics* **2014**, *196*, 853.

24. Phadke, S.S.; Maclean, C.J.; Zhao, S.Y.; Mueller, E.A.; Michelotti, L.A.; Norman, K.L.; Kumar, A.; James, T.Y. Genome-Wide Screen for Saccharomyces cerevisiae Genes Contributing to Opportunistic Pathogenicity in an Invertebrate Model Host. *G3 Genes|Genomes|Genetics* **2018**, *8*, 63.

25. Swinnen, S.; Schaerlaekens, K.; Pais, T.; Claesen, J.; Hubmann, G.; Yang, Y.; Demeke, M.; Foulquié-Moreno, M.R.; Goovaerts, A.; Souvereyns, K.; et al. Identification of novel causative genes determining the complex trait of high ethanol tolerance in yeast using pooled-segregant whole-genome sequence

analysis. *Genome Res.* **2012**, *22*, 975–984.

26. Wenger, J.W.; Schwartz, K.; Sherlock, G. Bulk Segregant Analysis by High-Throughput Sequencing Reveals a Novel Xylose Utilization Gene from Saccharomyces cerevisiae. *PLoS Genet.* **2010**, *6*, 18.

27. de Witt, R.N.; Kroukamp, H.; Van Zyl, W.H.; Paulsen, I.T.; Volschenk, H. QTL analysis of natural Saccharomyces cerevisiae isolates reveals unique alleles involved in lignocellulosic inhibitor tolerance. *FEMS Yeast Res.* **2019**, *19*.

28. Wang, Z.; Qi, Q.; Lin, Y.; Guo, Y.; Liu, Y.; Wang, Q. QTL analysis reveals genomic variants linked to high-temperature fermentation performance in the industrial yeast. *Biotechnol. Biofuels* **2019**, *12*.

29. Pomraning, K.R.; Smith, K.M.; Freitag, M. Bulk segregant analysis followed by high-throughput sequencing reveals the Neurospora cell cycle gene, ndc-1, to be allelic with the gene for ornithine decarboxylase, spe-1. *Eukaryot. Cell* **2011**, *10*, 724–733.

30. Daskalov, A.; Gladieux, P.; Heller, J.; Glass, N.L. Programmed Cell Death in Neurospora crassa Is Controlled by the Allorecognition Determinant rcd-1. *Genetics* **2019**, *213*, 1387–1400.

31. Kwon-Chung, K.J.; Edman, J.C.; Wickes, B.L. Genetic association of mating types and virulence in Cryptococcus neoformans. *Infect. Immun.* **1991**.

32. Loftus, B.J.; Fung, E.; Roncaglia, P.; Rowley, D.; Amedeo, P.; Bruno, D.; Vamathevan, J.; Miranda, M.; Anderson, I.J.; Fraser, J. a; et al. The genome of the basidiomycetous yeast and human pathogen Cryptococcus neoformans. *Science* **2005**, *307*, 1321–1324.

33. Kavanaugh, L.A.; Fraser, J.A.; Dietrich, F.S. Recent Evolution of the Human Pathogen Cryptococcus neoformans by Intervarietal Transfer of a 14-Gene Fragment. *Mol. Biol. Evol.* **2006**, *23*, 1879–1890.

34. Sun, S.; Xu, J. Chromosomal Rearrangements between Serotype A and D

Strains in Cryptococcus neoformans. *PLoS One* **2009**, *4*, e5524.

35. Priest, S.J.; Coelho, M.A.; Mixão, V.; Clancey, S.A.; Xu, Y.; Sun, S.; Gabaldón, T.; Heitman, J. Factors enforcing the species boundary between the human pathogens Cryptococcus neoformans and Cryptococcus deneoformans. *PLOS Genet.* **2021**, *17*, e1008871.

36. Jung, K.-W.; Yang, D.-H.; Maeng, S.; Lee, K.-T.; So, Y.-S.; Hong, J.; Choi, J.; Byun, H.-J.; Kim, H.; Bang, S.; et al. Systematic functional profiling of transcription factor networks in Cryptococcus neoformans. *Nat. Commun. 2015 61* **2015**, *6*, 1–14.

37. Lee, K.-T.; So, Y.-S.; Yang, D.-H.; Jung, K.-W.; Choi, J.; Lee, D.-G.; Kwon, H.; Jang, J.; Wang, L.L.; Cha, S.; et al. Systematic functional analysis of kinases in the fungal pathogen Cryptococcus neoformans. *Nat. Commun. 2016 71* **2016**, *7*, 1–16.

38. O'Meara, T.R.; Hay, C.; Price, M.S.; Giles, S.; Alspaugh, J.A. Cryptococcus neoformans histone acetyltransferase Gcn5 regulates fungal adaptation to the host. *Eukaryot. Cell* **2010**, *9*, 1193–1202.

39. Brandão, F.A.; Derengowski, L.S.; Albuquerque, P.; Nicola, A.M.; Silva-Pereira, I.; Poças-Fonseca, M.J. Histone deacetylases inhibitors effects on Cryptococcus neoformans major virulence phenotypes. *https://doi.org/10.1080/21505594.2015.1038014* **2015**, *6*, 618–630.

40. Garcia, A.W.A.; Kinskovski, U.P.; Diehl, C.; Reuwsaat, J.C.V.; Motta de Souza, H.; Pinto, H.B.; Trentin, D. da S.; de Oliveira, H.C.; Rodrigues, M.L.; Becker, E.M.; et al. Participation of Zip3, a ZIP domain-containing protein, in stress response and virulence in Cryptococcus gattii. *Fungal Genet. Biol.* **2020**, *144*, 103438.

41. Squizani, E.D.; Reuwsaat, J.C.V.; Motta, H.; Tavanti, A.; Kmetzsch, L. Calcium: a central player in Cryptococcus biology. *Fungal Biol. Rev.* **2021**, *36*, 27–41.

42. Erickson, T.; Liu, L.; Gueyikian, A.; Zhu, X.; Gibbons, J.; Williamson, P.R.

Multiple virulence factors of Cryptococcus neoformans are dependent on VPH1. *Mol. Microbiol.* **2001**, *42*, 1121–1131.

43. Haas, R.; Horev, G.; Lipkin, E.; Kesten, I.; Portnoy, M.; Buhnik-Rosenblau, K.; Soller, M.; Kashi, Y. Mapping Ethanol Tolerance in Budding Yeast Reveals High Genetic Variation in a Wild Isolate. *Front. Genet.* **2019**, *10*.

44. Kwon-Chung, K.J.; Edman, J.C.; Wickes, B.L. Genetic association of mating types and virulence in Cryptococcus neoformans. *Infect. Immun.* **1991**, *60*, 602–605.

45. Samarasinghe, H.; Aceituno-Caicedo, D.; Cogliati, M.; Kwon-Chung, K.J.; Rickerts, V.; Velegraki, A.; Ackaglar, S.; Xu, J. Genetic Factors and Genotype-Environment Interactions Contribute to Variation in Melanin Production in the Fungal Pathogen Cryptococcus neoformans. *Sci. Rep.* **2018**, *8*.

46. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data Available online: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on Jul 7, 2021).

47. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120.

48. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760.

49. Koboldt, D.C.; Chen, K.; Wylie, T.; Larson, D.E.; McLellan, M.D.; Mardis, E.R.; Weinstock, G.M.; Wilson, R.K.; Ding, L. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **2009**, *25*, 2283.

50. GA, V. der A.; MO, C.; C, H.; R, P.; G, D.A.; A, L.-M.; T, J.; K, S.; D, R.; J, T.; et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **2013**, *43*.

51. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land,

S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin).* **2012**, *6*, 80.

52. Gautier, M.; Vitalis, R.; Flori, L.; Estoup, A. f-statistics estimation and admixture graph construction with Pool-Seq or allele count data using the R package poolfstat. *bioRxiv* **2021**, 2021.05.28.445945.

53. Fan, Y.; Lin, X. Multiple Applications of a Transient CRISPR-Cas9 Coupled with Electroporation (TRACE) System in the Cryptococcus neoformans Species Complex. *Genetics* **2018**, *208*, 1357–1372.

54. Fang, Y.; Cui, L.; Gu, B.; Arredondo, F.; Tyler, B.M. Efficient Genome Editing in the Oomycete Phytophthora sojae Using CRISPR/Cas9. *Curr. Protoc. Microbiol.* **2017**, *44*, 21A.1.1-21A.1.26.

55. Fu, C.; Davy, A.; Holmes, S.; Sun, S.; Yadav, V.; Gusa, A.; Coelho, M.A.; Heitman, J. Dynamic genome plasticity during unisexual reproduction in the human fungal pathogen Cryptococcus deneoformans. *bioRxiv* **2021**, 2021.06.01.446667.

Chapter 5

Hybrids and Hybridization in the *Cryptococcus neoformans* and *Cryptococcus gattii* Species Complexes

5.1 Preface

The pathogenic species of the *Cryptococcus* genus include *C. neoformans*, *C. deneoformans* and *C. gattii*. The three species differ in cell capsule structure, geographical distribution, pathogenicity, and host range. Ongoing hybridization between the three species in natural environments is evidenced by the presence of hybrid strains in environmental sources and the increasing prevalence of infections caused by hybrids in clinical settings. In this literature review, we synthesized the current state of knowledge on the global epidemiology of cryptococcal hybrids and identified them as a model system for studying hybridization among yeast species. We identified knowledge gaps in cryptococcal hybrid research and proposed experimental frameworks that could address them.

There is an ongoing debate on whether *C. neoformans* and *C. deneoformans* should be considered separate species. This review paper uses an alternative nomenclature system where *C. neoformans* and *C. deneoformans* are designated as varieties of a single species complex. This manuscript is published in Infection, Genetics and Evolution volume 66, pages 245-255. I am the primary contributor of this work. I conducted the literature review, analyses and drafting of the manuscript. Dr. Jianping Xu is a co-author and made significant contributions to the manuscript.

5.2 Abstract

The basidiomycetous yeasts of the *Cryptococcus neoformans* and *Cryptococcus gattii* species complexes (CNSC and CGSC respectively) are the causative agents of cryptococcosis, a set of life-threatening diseases affecting the central nervous system, lungs, skin, and other body sites of humans and other mammals. Both the CNSC and CGSC can be subdivided into varieties, serotypes, molecular types, and lineages based on structural variations, molecular characteristics and genetic sequences. Hybridization between the haploid lineages within and between the two species complexes is known to occur in natural and clinical settings, giving rise to intraspecific and interspecific diploid/aneuploid hybrid strains. Since their initial discovery in 1977, cryptococcal hybrids have been increasingly discovered in both clinical and environmental settings with over 30% of all cryptococcal infections in some regions of Europe being caused by hybrid strains. This review summarizes the major findings to date on cryptococcal hybrids, including their possible origins, prevalence, genomic profiles and phenotypic characteristics. Our analyses suggest that CNSC and CGSC can be an excellent model system for studying fungal hybridization.

5.3 Introduction

Biological hybridization refers to sexual mating between two organisms belonging to genetically different populations, varieties, and species. Hybridization is a common occurrence in many species of plants, animals, and eukaryotic microbes such as fungi. The genotypic and phenotypic consequences of hybridization are often unpredictable with possible outcomes ranging from hybrid inviability to hybrid vigor. Hybrid offspring can yield novel information about the genetic bases of phenotypic differences between their parents and the evolutionary trajectory of the parental populations. In addition to being an interesting and widespread phenomenon from an evolutionary perspective, hybridization and hybrids can also have applied significance. For example, commercial agricultural practices have been increasingly adopting hybrid cultivars due to their higher yields and other favorable traits (reviewed in Goulet et al., 2017). Similarly, some of the deadly pathogens responsible for outbreaks were generated through hybridization events, including strains of the human parasite *Toxoplasma gondii* (Wendte et al., 2010) and the causative strain of

H1N1 influenza pandemic in 2009 (Christman et al., 2011). *Saccharomyces pastorianus*, the yeast used in lager beer production, was shown to be an interspecific hybrid between *Saccharomyces cerevisiae* and *Saccharomyces eubayanus* (Krogerus et al., 2017; Libkind et al., 2011).

The closely related, basidiomycetous yeasts of the *Cryptococcus neoformans* and *Cryptococcus gattii* species complexes (CNSC and CGSC) are the causative agents of life-threatening cryptococcosis in humans and other mammals. Hybridization is known to occur between haploid isolates of different lineages within the same species complex (in this review called intraspecific hybrids) as well as between isolates belonging to the two sister species complexes (interspecific hybrids), giving rise to diploid/aneuploid hybrid strains. While *C. neoformans* was first described in 1894 (Busse, 1894; Sanfelice, 1894), hybrid cryptococcal strains were not discovered until the late 1970s (Bennett et al., 1977). Since then, significant research efforts have focused on elucidating the origin, genetic composition and phenotypic patterns of these hybrids, especially in the light of their increasing prevalence among clinical samples in certain European regions. Furthermore, the fully-sequenced genomes, well-developed morphological and phenotypic assays, and ease of genetic manipulation make the human pathogenic *Cryptococcus* an ideal model system for studying hybridization. This review aims to describe the current body of knowledge on cryptococcal hybrids including possible origins, mechanisms for subsequent evolution and their genetic and phenotypic profiles in comparison to the parental populations.

5.4 *C. neoformans* and *C. gattii* species complexes

The basidiomycetous yeasts of the CNSC and CGSC are the causative agents of cryptococcosis, a set of life-threatening diseases affecting the central nervous system, skin, lungs, and other body sites in humans and other mammals (Charlier et al., 2009; Goldman et al., 1994; Husain et al., 2001). Cryptococcal infections are among the most common fungal infections of the central nervous system, clinically presenting as inflammation of the meninges leading to meningoencephalitis (Del Valle and Pina-Oviedo, 2006). Each year, approximately 220 000 cases of cryptococcal meningitis are diagnosed worldwide with 180 000 of them resulting in mortality (Rajasingham et al., 2017). While naturally found in soil, pigeon droppings and tree barks, *C.*

neoformans and *C. gattii* can enter hosts via the respiratory tract upon inhalation of the spores (Velagapudi et al., 2009). Once established within the alveoli of the lungs, cryptococcal cells can disseminate into the bloodstream resulting in infections at secondary sites. They can cross the blood-brain barrier to infect the meninges, the protective three-layered membrane that encloses the brain and spinal cord, leading to fatal meningoencephalitis (Stie and Fox, 2012; Vu et al., 2014). The host immune system combats the invading yeast cells by incorporating them into macrophages through phagocytosis inside which the yeast cells are bombarded with oxygen and nitrogen radicals that could cause fatal cellular damage (Forman and Torres, 2002; Nathan and Shiloh, 2000). Cryptococcal pathogens have evolved mechanisms to counteract these host immune responses, leading to their persistence as an intracellular parasite within mammalian macrophages (Alvarez and Casadevall, 2007; Feldmesser et al., 2000; Ma et al., 2007; Voelz et al., 2009). If left untreated, cryptococcal infections will be fatal to the host.

Extensive genetic diversity exists among isolates in environmental and clinical populations of CNSC and CGSC with various molecular techniques such as amplified fragment length polymorphisms (AFLP), PCR fingerprinting, and multilocus sequence typing being routinely used to identify the different genotypes (Boekhout et al., 2001; Cogliati, 2013; Fraser et al., 2005; Litvintseva et al., 2006; Meyer et al., 1993, 1999, 2003, 2009; Taylor and Fisher, 2003; Viviani et al., 1997). Cryptococcal strains are also categorized into four main serotypes named A, B, C and D based on the structural variations in glucuronoxylomannan (G_AXM), the most abundant polysaccharide in the cryptococcal capsule (Bhattacharjee et al., 1984). The serotypes and molecular genotypes show broad correlations (**Table 1; Figure 1**). Strains of serotype A belong to one of three molecular types (VNI, VNII and VNB) and three AFLP genotypes (AFLP1, AFLP1A and AFLP1B). Serotype D strains correspond to molecular types VNIV and AFLP2. Hybrids derived from mating between serotype A and D isolates, termed AD hybrids, are assigned molecular types VNIII and AFLP3. Together, serotypes A, D and AD hybrids make up the *Cryptococcus neoformans* species complex. Serotypes A and D have been elevated to variety status, with serotype A designated *C. neoformans* var. *grubii* and serotype D designated *C. neoformans* var.

neoformans (Franzot et al., 1999). Serotypes B and C do not strictly correspond to separate molecular types but seem to share the four VG molecular types, VGI to VGIV, and five AFLP genotypes (AFLP4, AFLP5, AFLP6, AFLP7, AFLP10): together, they form the *Cryptococcus gattii* species complex. Unlike A and D, serotypes B and C have not been granted variety status (Kwon-Chung et al., 2017). Phylogenetic analyses based on nuclear and mitochondrial gene sequences support the existence of these distinct lineages within CNSC and CGSC (Hagen et al., 2015, reviewed in Xu et al., 2011). There is an ongoing debate as to whether each lineage should be designated a distinct species or maintained as different varieties and molecular types within the larger species complexes (Hagen et al., 2017, 2015; Kwon-Chung et al., 2017; Kwon-Chung and Varma, 2006). While we use the species complex nomenclature in this paper, we have provided the proposed, alternative names for the different lineages in **Table 1** to help readers keep track of the historical and ongoing developments.

Table 1: Recently proposed names for the *C. neoformans* and the *C. gattii* species complexes

Current name	Serotype(s)	Molecular type(s)	Proposed new species name
C. neoformans var. *grubii*	A	VNI/VNII/VNB (AFLP1, AFLP1A, AFLP1B)	*C. neoformans*
var. *neoformans*	D	VNIV (AFLP2)	*C. deneoformans*
AD hybrid	A, D, or AD	VNIII (AFLP3)	*C.neoformans* × *C. deneoformans* hybrid
C. gattii	B, C	VGI (AFLP4)	*C. gattii*
		VGII (AFLP6)	*C. deuterogattii*
		VGIII (AFLP5)	*C. bacillisporus*
		VGIV (AFLP7)	*C. tetragattii*
		VGIV/VGIIIc (AFLP10)	*C. decagattii*
BD hybrid	BD	AFLP8	*C. deneoformans* × *C. gattii* hybrid
AB hybrid	AB	AFLP9	*C. neoformans* × *C. gattii* hybrid
AB hybrid	AB	AFLP11	*C. neoformans* × *C. deuterogattii* hybrid

*adapted from (Hagen et al., 2015; 2017; Kwon-Chung et al., 2017)

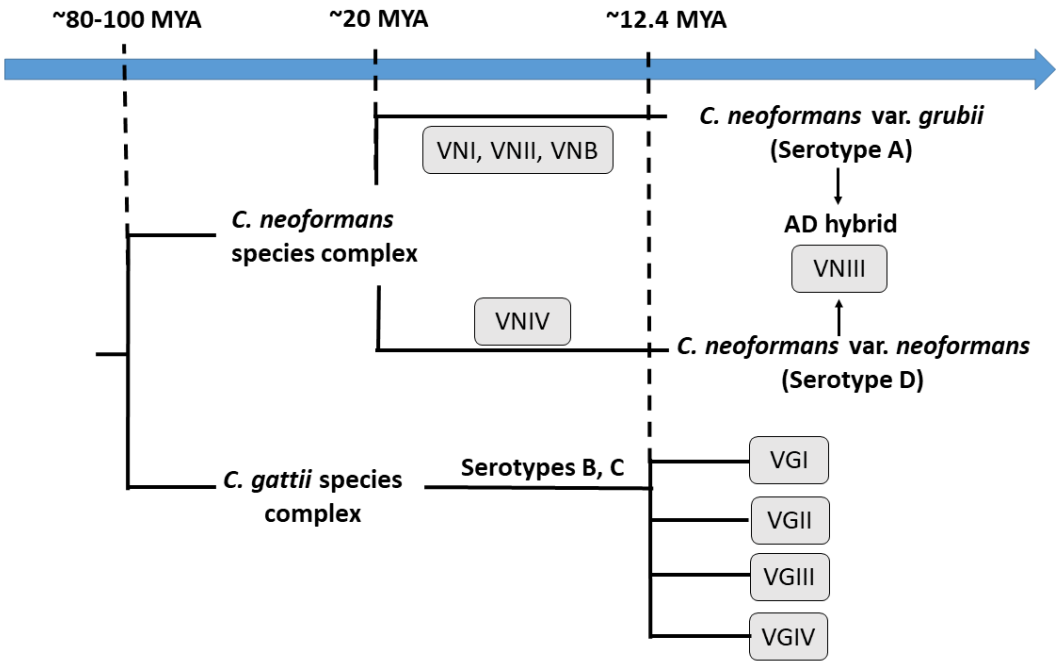


Figure 1: Estimated divergence times between the cryptococcal lineages. C. *neoformans* species complex (CNSC) and the C. *gattii* species complex (CGSC) include several, diverged lineages/molecular types whose phylogenetic relationships are supported by sequence analyses. CNSC includes serotype A (molecular types VNI, VNII, VNB), serotype D (VNIV) and AD hybrids (VNIII). CGSC includes serotypes B and C and molecular types VGI to VGIV. The split between CNSC and CGSC occurred approximately ~80-100 million years ago. Serotypes A and D separated ~20 MYA while the CGSC lineages diverged from each other approximately ~12 MYA (Casadevall et al., 2017; D'Souza et al., 2011; Sharpton et al., 2008; Xu et al., 2000).

Up until the early 2000s, *C. gattii* was considered a variety of CNSC and was referred to as *C. neoformans* var. *gattii*. Accumulation of empirical data over the years showing the divergence of CGSC strains from CNSC in DNA sequence, morphology, ecology, epidemiology, and virulence led to it being recognized as a separate species in 2002 (Kwon-Chung et al., 2002). Indeed, the sister species complexes CGSC and CNSC show characteristic differences in geographical distribution, morphology and virulence. CNSC is primarily an opportunistic pathogen that infects immunocompromised individuals although infections in apparently healthy individuals have been reported (Arsenijevic et al., 2014; Chen et al., 2008; Favalessa et al., 2014; M. Li et al., 2012). In contrast, CGSC strains belonging to the prevalent VGI and VGII lineages are able to cause infections primarily in immunocompetent hosts (Chen et al., 2014, 2012; Phillips et al., 2015). CGSC strains are often found on decaying tree barks whereas CNSC is typically associated primarily with pigeon

excreta and soil, and sometimes trees (Cogliati et al., 2016; Mitchell et al., 2011). CGSC was originally thought to be restricted to tropical and subtropical regions until in 2001 a major cryptococcal outbreak caused by a recombinant strain of CGSC emerged in Vancouver Island in British Columbia, Canada (Kidd et al., 2004; Kwon-Chung and Bennett, 1984; Sorrell, 2001). CGSC natural isolates and clinical cases have since been reported in Pacific Northwest of the US, other regions in North America, South America, Europe and Asia (Cogliati, 2013). Within CNSC, *C. neoformans* var. *grubii* has a worldwide distribution with the majority of all reported clinical and environmental isolates in Europe, Asia and Africa (59%, 81% and 68% respectively) belonging to this variety (Cogliati, 2013). Given the tropical climate, CGSC isolates are predominant in Oceania countries (39%) followed by *C. neoformans* var. *grubii* at 27% (Cogliati, 2013). *C. neoformans* var. *neoformans* also has a global distribution but is more commonly found in temperate climates, especially in Europe where it is responsible for ~30% of all reported cryptococcal infections (Arsenijevic et al., 2014; Viviani et al., 2006).

5.5 Sexual cycle of CNSC and CGSC

C. neoformans and *C. gatii* have a heterothallic mating system with haploid strains belonging to one of two mating types as determined by the mating type locus (*MAT*) located on chromosome 4 (Lengeler et al., 2002). When fertile *MAT***a** and *MATα* strains come into contact with each other, the cells fuse and form a dikaryotic hypha inside which the two haploid nuclei co-exist within each cell (Kwon-Chung, 1976a, 1976b, 1975). At the terminus of each hypha, the cells enlarge forming a basidium within which the two haploid nuclei fuse to form a diploid nucleus. This is then followed by one round of DNA replication and meiosis to produce four, recombinant, haploid nuclei (Idnurm, 2010). These nuclei then undergo multiple rounds of mitosis with each haploid progeny nucleus packaged into one basidiospore. Each basidium typically has four chains of basidiospores. These sexual spores are then released into the environment and can infect humans as well as other hosts (Velagapudi et al., 2009).

The prevalence of the two *MAT* alleles varies among wild populations of CNSC and CGSC. *MAT***a** allele is rare in natural populations of CNSC with less than 1% of

strains in most populations belonging to this mating type (Chowdhary et al., 2011; Feng et al., 2008; Freire et al., 2012; Keller et al., 2003; Lengeler et al., 2000; Sanchini et al., 2014; Viviani et al., 2003; Yan et al., 2002). Therefore, the global CNSC populations were thought to be clonal until evidence of sexual recombination was found in the form of incongruent gene genealogies and linkage equilibrium between loci (Hiremath et al., 2008; Litvintseva et al., 2005; Xu et al., 2000; Xu and Mitchell, 2003). Later, environmental VNI strains of opposite mating types isolated from similar areas in Spain were shown to be fertile and capable of mating with each other in the laboratory, suggesting that CNSC populations can undergo sexual recombination in nature (Cogliati et al., 2016). Furthermore, a unique serotype A population was discovered in Botswana where 25% of the population was carrying the rare *MAT***a** allele (Litvintseva et al., 2003). This population was given the molecular type VNB and was determined to be highly recombinant. Originally thought to be endemic to Botswanan, VNB strains have since been isolated from other regions in South Africa and South America (Litvintseva et al., 2011; Rhodes et al., 2017). Recombinant genotypes in nature could also be generated through same-sex mating between *MATα* strains (Lin et al., 2005). Same-sex mating has been demonstrated in the lab while phylogenetic incompatibility, a signature of recombination, has been reported from populations composed exclusively of *MATα* strains (Hiremath et al., 2008; Lin et al., 2005; Ni et al., 2013). In CGSC, the *MAT***a**:*MATα* ratio can vary significantly between populations depending on the molecular type and geographical origin of the isolates. One study found VGI strains isolated from patients in China to be exclusively *MATα* (Feng et al., 2008) while another study found the prevalence of *MAT***a** among clinical and environmental VGI isolates in Apulia, Italy to be 95% (Montagna et al., 2018). In contrast, *MATα* is more prevalent in VGIII populations with the proportion of *MAT***a** strains reported to be as low as 0%-10% in some populations (Escandon et al., 2006; Lockhart et al., 2013; Hagen et al. 2015).

Sexual reproduction can occur between different molecular types within each species complex as well as between strains of CNSC and CGSC, giving rise to intraspecific and interspecific cryptococcal hybrids respectively. Analysis of the mating type pheromone gene *MFα* in 122 strains encompassing *C. gattii, C. neoformans* var.

grubii and *C. neoformans* var. *neoformans* revealed single nucleotide polymorphisms, chromosomal locations and gene copy numbers unique to each lineage (Chaturvedi et al., 2002). However, despite these differences, selected *MATα* strains of serotypes A, B, C and D were able to mate with *MAT***a** tester strains of *C. neoformans* var. *neoformans* (NIH430 and NIH433), suggesting that there is limited pre-zygotic reproductive isolation among cryptococcal lineages.

5.6 Intraspecific hybrids

CNSC hybrids

AD hybrids (VNIII, AFLP3), arising from hybridization between serotype A (VNI, VNII, VNB) and serotype D (VNIV), were the first cryptococcal hybrids to be identified. In an epidemiological survey conducted in the United States in 1977, 11 out of the 272 tested cryptococcal isolates consistently reacted with both A and D antisera and were therefore typed as AD hybrids (Bennett et al., 1977). Since then, AD hybrids who are assigned the molecular type VNIII have been isolated from environmental and clinical sources in many countries (**Table 2**). In an interesting twist, the original *C. neoformans* isolate discovered in peach juice from Italy in 1894 which was thought to be of serotype D was recently determined to be an AD hybrid (Boekhout et al., 2001; Sanfelice, 1894; Viviani et al., 1997). In fact, AD hybrids are the most common of all cryptococcal hybrids: globally, AD hybrids show the highest prevalence in Europe at 18%, followed by the United Sates at 6% (Cogliati, 2013).

Most AD hybrids are heterozygous and diploid at the *MAT* locus with each allele likely to have originated from one parent (Cogliati et al., 2001). The majority of the AD hybrids identified to date contain *MATα* of the A parent (Aα) and *MAT**a*** of the D parent (D**a**), forming αAD**a** hybrids (Cogliati et al., 2006; Desnos-Ollivier et al., 2015). Albeit less common, hybrids with **a**AD*α* mating type combination have also been reported (Arsenijevic et al., 2014; Lengeler et al., 2001; Litvintseva et al., 2007; Yan et al., 2002). AD hybrids generated from same-sex mating in nature as well as in laboratory have also been reported. For example, in 2007, three environmental isolates collected from pigeon excreta in North Carolina, USA were determined to be αADα hybrids (Lin et al., 2007). The researchers were then able to construct an αADα hybrid in the laboratory using H99 (Aα) and JEC21 (Dα). Same-sex AD hybrids have since

been reported in Italy (W. Li et al., 2012) and Libya (Ellabib et al., 2016). In addition to hybrids generated from strains of different serotypes, hybrids produced from mating of strains belonging to the same serotype (but different molecular types or lineages) have also been detected in natural cryptococcal populations. For example, a clinical isolate from Apulia, Italy was determined to be an αDDα hybrid (Montagna et al., 2018). A survey of 489 clinical and environmental isolates of CNSC encompassing strains from six continents (Asia, Africa, Australia, Europe, and North and South Americas) revealed 19 αAAα hybrids (Lin et al., 2009). Also, an environment isolate found in the soil from Tripoli, Libya was determined to be an αAAα hybrid (Ellabib et al., 2016).

Most **a**ADα and αADα strains seem to share a unique mating-type-specific cell identity determinant *SXI1α* allele that bears a truncation at the C-terminus which seems to have enhanced the fertility of the Dα parent of the hybrids (Lin et al., 2007). This allele is present in the global natural Dα population at a relatively low level of ~13%. The enhanced fertility conferred by this novel *SXI1α* allele likely promoted mating between serotypes A and D explaining the high prevalence of AD hybrids in cryptococcal populations.

It is possible that some of the current haploid CNSC strains are themselves a product of ancient hybridization events between CNSC lineages that were haploidized during subsequent evolution and somatic recombination. In 2017, a population genomics analyses of 188 VNI, VNII, and VNB isolates found evidence of ancestral hybridization events between different VN lineages: these events would explain the current cryptococcal strains having certain regions within single chromosomes derived from different ancestors (Rhodes et al., 2017). This suggested that hybridization between the lineages VNI, VNII and VNB has been occurring throughout the history of CNSC and that some of the current haploid populations are likely recombinant products of these historical hybridization events.

Table 2: Prevalence of AD hybrids in different countries and regions in the world.

Country	Source	Percentage of hybrid isolates	# hybrid isolates / # of total cryptococcal isolates	Year(s) when sample was obtained	Reference
Europe					
Apulia, Southern Italy	Environmental	1	1/95	2018	(Montagna et al., 2018)
Slovenia	Clinical	2.2	1/46	2017	(Tomazin et al., 2017)
Italy	Clinical	12	21/170	2017	(Pini et al., 2017)
	Environmental	0	0/32		
Spain	Clinical	25	7/28	2016	(Gago et al., 2017)
Mediterranean basin	Environmental (trees)	7.4	35/474	2016	(Cogliati et al., 2016)
Denmark	Both	12	13/108	2016	(Hagen et al., 2016)
Portugal	Clinical	30.9	38/122	2014	(Maduro et al., 2014)
Serbia	Clinical	8.8	3/34	2014	(Arsenijevic et al., 2014)
Portugal	Environmental	14.3	4/28	2013	(Ferreira et al., 2014)
Croatia	Clinical	6.25	3/48	2011	(Mlinaric-Missoni et al., 2011)
Europe	Clinical	19	N/A	2006	(Viviani et al., 2006)
Italy	Clinical	31	N/A	2002	(Viviani et al., 2002)
Italy	Clinical	36.8	49/133	2001	(Cogliati et al., 2001)
France	Clinical	20	80/400	1997-2001	(Desnos-Ollivier et al., 2015)
Germany	Clinical	6	6/104	2004-2010	(Sanchini et al., 2014)
Netherland	Clinical	3.6	11/300	1977-2007	(Hagen et al., 2012)
Asia					
China	Clinical	1.8	2/109	2010	(Chen et al., 2010)
China	Clinical	1.7	2/115	2008	(Feng et al., 2008)
Turkey	Environmental	3.8	1/26	2006	(Saracli et al., 2006)
Thailand	Clinical	1.8	3/169	1996	(Sukroongreung et al., 1996)
Japan	Clinical/Environmental	3	1/33	1983	(Hironaga et al., 1983)
Africa					
Ivory Coast	Clinical	11	40/363	2016	(Kassi et al., 2016)
South Africa	Clinical	4	7/177	2014	(Van Wyk et al., 2014)
South Africa	Pediatric clinical	1.2	1/82	2011	(Miglia et al., 2011)
Malawi	Clinical	6.7	1/15	2005	(Litvintseva et al., 2005)
North America					
North Carolina, USA	Environmental	7.1	54/762	2005	(Litvintseva et al., 2005)
	Clinical	2.4	1/42		
New York, USA	Clinical	2.5	1/39	2000	(Steenbergen and Casadevall, 2000)
South America					
Iberoamerican countries	Clinical	7.4	13/266	2003	(Meyer et al., 2003)
	Environmental	1.5	1/67		
Brazil	Environmental	7.4	2/27	2004	(Oliveira et al., 2004)
	Clinical	6.5	2/31		
Brazil	Environmental	1.3	1/80	1987-1998	(Nishikawa et al., 2003)
	Clinical	1.3	5/387		
Rio de Janeiro, Brazil	Environmental	1.2	1/83	2006	(Baroni et al., 2006)
Australia					
Australia	Clinical	1.6	1/60	1999	(Meyer et al., 1999)

CGSC hybrids

Under mating-inducing conditions, strains of CGSC belonging to serotypes B and C have been found capable of mating with each other in the laboratory (Campbell et al., 2005; Fraser et al., 2003; Kidd et al., 2004). A study of 120 strains belonging to serotype B and molecular types VGI, VGII and VGIII from Australia found most of the isolates to be sterile while the fertile, robust maters tended to be of molecular type VGII and *MATα* (Campbell et al., 2005). All nine robust maters produced basidiospores when mated with JEC20 (D**a**, VNIV) and B4546 (C**a**, VGIII) suggesting that most instances of hybridization within CGSC likely involve VGII strains as a partner. A later study found spore germination rate to be significantly low (< 1%) in laboratory-derived reciprocal crosses of VGII (*MATα)* x VGIII (*MAT**a***) and VGII (*MAT**a)*** x VGIII (*MATα*) (Voelz et al., 2013). From the spores that successfully germinated in the two crosses, 18/18 and 9/16 were diploid/aneuploid respectively. However, when strains of serotypes B and C mated with each other, they produced proper dikaryotic hyphae with clamp connections that were morphologically similar to those produced when CGSC strains of the same serotype mated with each other, suggesting that some degree of post-zygotic reproductive isolation exists among CGSC lineages (Campbell et al., 2005; Kwon-Chung, 1976b) . Compared to AD hybrids, intraspecific CGSC hybrids occur much less frequently and to the best of our knowledge, no hybrid CGSC strains have been isolated from natural sources.

5.7 Interspecific hybrids

Hybridization between the two cryptococcal species complexes seems to be a much rarer occurrence in nature compared to hybridization between CNSC lineages. There have been several reported cases of AB and BD hybrids causing infections in patients, although interspecific cryptococcal hybrids have never been isolated from environmental sources. The first naturally occurring CNSC x CGSC hybrid was reported in 2006 when three strains isolated from the cerebrospinal fluid (CSF) of two patients in the Netherlands were determined to be diploids of serotype BD (Bovers et al., 2006). They carried two *MAT* alleles with *MAT**a*** originating from the D parent and *MATα* from the B parent (**a**BDα). Based on these findings, they estimated that BD hybrids make up about 1% of the clinical isolates found in Netherlands between 1977

and 2007. The first natural serotype AB strain was reported in 2008 when a strain isolated from an AIDS patient in Canada was determined to be an AB hybrid with the *MATα* allele originating from the serotype A parent (Bovers et al., 2008). Since ***MATa*** was not detected in this strain, it was hypothesized that it originated from a hybridization between an Aα strain and a B**a** strain and later lost the ***MATa*** allele. In 2012, four αAB**a** hybrids with both *MAT* alleles intact were identified among clinical samples obtained from the CSF of patients in Brazil (n=2), Columbia (n=1) and India (n=1) (Aminnejad et al., 2012). Interspecific cyptococcal hybrids have since been reported in Germany (Smith et al., 2015), Denmark (Hagen et al., 2016) and the United States (Rhodes et al., 2017).

So far, the recovery of interspecific hybrids has been limited to clinical sources: to the best of our knowledge, none have been isolated from the environment. Strains of CNSC and CGSC are known to co-occur in the same geographical regions (Chau et al., 2010; Choi et al., 2010; Chowdhary et al., 2012; Cogliati et al., 2016; Kaocharoen et al., 2013) and lack of hybridization may indicate reproductive isolation between the two groups in nature. While it was estimated that CNSC and CGSC diverged from each other up to ~100 million years ago (Sharpton et al., 2008) following the breakup of the supercontinent Pangea, human migration and associated anthropomorphic activity have brought these groups from previous geographical isolation into close contact (reviewed in Casadevall et al., 2017). The rarity of interspecific cryptococcal hybrids and their apparent absence in environmental sources are consistent with extensive divergence and reproductive isolation between the two sister species complexes. However, in the laboratory, mating between strains of CNSC and CGSC can be induced, consistent with the lack of an absolute pre-zygotic reproductive barrier (Fraser et al., 2003). It should be noted that when strains of serotypes A or D mate with strains of serotypes B and C, they often fail to complete the sexual cycle, indicative of a strong post-zygotic reproductive barrier (Kwon-Chung, 1976b). Together, these observations suggest that while the two species are still genetically and phenotypically compatible enough to initiate mating with each other, they may not produce viable hybrid haploid progeny, leading to their rare occurrence in

environmental and clinical samples. Up to now, there have been no reports of interspecific hybrids of serotypes AC and CD.

5.8 Origins of cryptococcal hybrids

Several studies have investigated the origins of cryptococcal hybrids found in environmental and clinical settings. In 2000, Xu et al. conducted a multiple gene genealogy analysis to determine if AD hybrids originated from a recent vs. ancient hybridization event between different cryptococcal serotypes (Xu et al., 2000). They sequenced four genes (*ITS, LAC1, URA5* and mitochondrial *LnRNA)* in a set of 34 cryptococcal strains encompassing serotypes A, B, C, D and five AD hybrids. Phylogenetic analyses of the four gene sequences showed that the individual alleles of AD hybrids unambiguously clustered with either serotype A or serotype D sequences. Furthermore, the five AD hybrids in this study were not identical to each other. This was consistent with the AD strains arising from more than one hybridization event between strains of serotypes A and D. Since they did not find novel gene lineages which would be consistent with ancient hybridization events, they concluded that the hybrids likely arose from multiple hybridization events made possible by recent global dispersal of previously geographically isolated cryptococcal lineages (**Figure 2**). In a follow up study where they analyzed the *LAC1* gene genealogy in 14 AD strains, Xu et al. confirmed that there had been at least three independent hybridization events to yield those 14 AD hybrid strains (Xu et al., 2002).

While confirming the above findings, a gene genealogy analysis conducted by Litvintseva et al. in 2007 further revealed that many of the global AD hybrids may have originated in Africa (Litvintseva et al., 2007). The presence of the rare *MAT**a*** allele in approximately half of the globally isolated AD strains, either in the form of **a**ADα or αAD**a** (W. Li et al., 2012; Lin et al., 2005, 2008; Yan et al., 2002) raised the possibility that cryptococcal hybrids with A**a** may be descended from ancient hybridization events when *MAT**a*** was more prevalent in the population. Based on the genealogies of *CAP10, URE1* and *GPD1*, Litvintseva et al. determined that the three **a**ADα strains examined in this study were closely related to VNB strains in which *MAT**a*** prevalence is unusually high at 25% (Litvintseva et al., 2003). Furthermore, they found evidence of hybrid vigor as the hybrid strains were more resistant to UV

irradiation than Botswanan serotype A strains. Their results are consistent with **a**ADα originating in Africa from recent hybridization events between VNB strains and serotype D strains followed by global dispersal, presumably facilitated by their higher fitness (**Figure 2**).

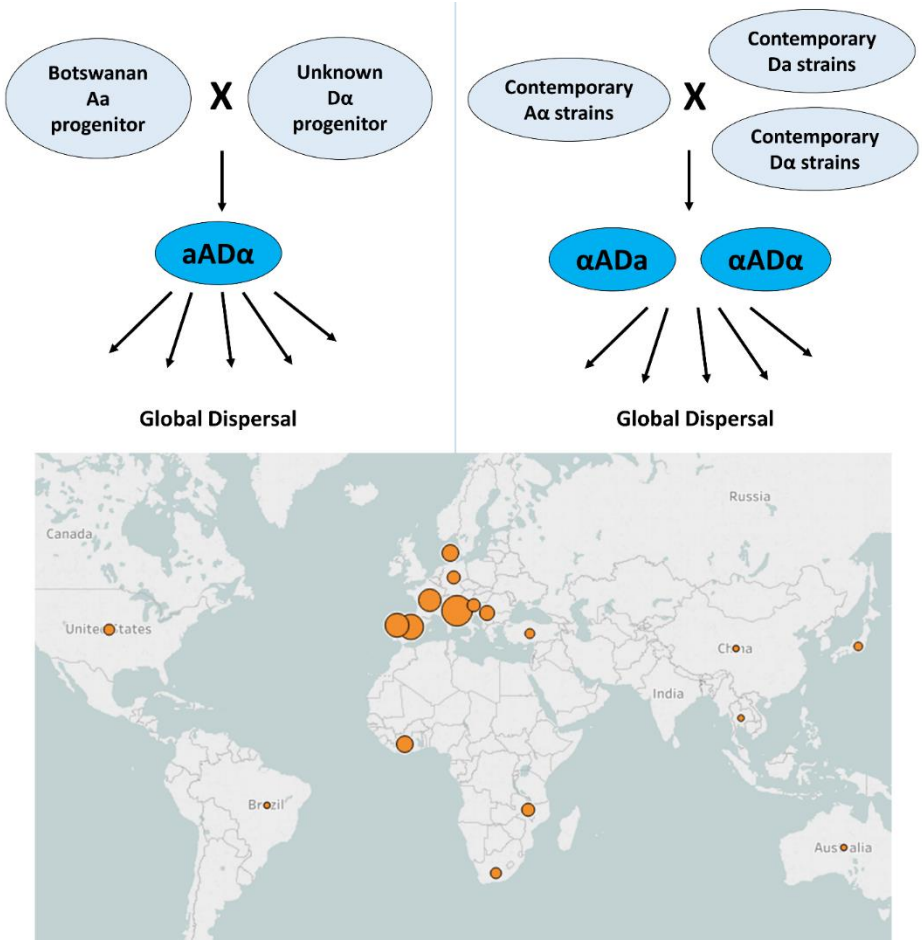


Figure 2: Origin and dispersal of the current global AD population. aADα hybrids are believed to have arisen in Africa from multiple, recent hybridization events between the VNB (Aa) strains from Botswana and an unknown Dα progenitor. αADa and αADα hybrids likely originated from multiple, recent hybridization events between contemporary Aα strains and Da/Dα strains. The origins of these AD hybrids were followed by global dispersal resulting in the current global AD population (Li et al., 2012; Litvintseva et al., 2007; Xu et al., 2000; 2002). The orange circles represent the countries in which AD hybrids have been reported: the size of the circles is proportional to the prevalence of AD hybrids in these countries. The prevalence statistics are based on the references found in Table 2.

Multilocus sequence typing and comparative genome hybridization of 31 AD hybrids further supported recent, multiple hybridization events leading to the emergence of the current global AD population (Li et al., 2012). While individual serotype A and D

alleles of the AD hybrids clustered with those of haploid A and D strains in this phylogenetic analysis, they were not identical to the currently sequenced haploid alleles. This and other results suggested that there is still undiscovered genetic diversity in haploid serotype A and D strains in nature (Xu et al. 2002; Xu and Mitchell 2003). Alternatively, the results are also consistent with AD hybrids experiencing mutation accumulation and ongoing evolutionary divergence independent of their parent lineages. Similar studies to investigate the origin of interspecific cryptococcal hybrids have not been conducted, primarily due to their rare occurrence. Given that all reported interspecific cryptococcal hybrids to date are of clinical origin and their hosts were known to have traveled to different countries prior to showing symptoms, it is possible that they were co-infected with strains of the two species and they hybridized within the host during the infection.

5.9 Hybrid genomes are highly variable and dynamic

Unlike their haploid parents, intraspecific and interspecific cryptococcal hybrids are often diploid or aneuploid (Aminnejad et al., 2012; Bovers et al., 2006, 2008; Hu et al., 2008; Lengeler et al., 2001; W. Li et al., 2012; Lin et al., 2007; Tanaka et al., 1999; Vogan et al., 2013). Cryptococcal hybrids are usually heterozygous at multiple loci including the mating locus (Lengeler et al., 2001; Lin et al., 2007; Sun and Xu, 2007; Xu and Mitchell 2003). Aneuploidy of cryptococcal hybrids is a result of impaired meiosis and mitotic recombination that occur during hybridization as well as subsequent vegetative growth. The genomes of serotypes A and D have over 30 chromosomal rearrangements and show 10-15% sequence divergence (Kavanaugh et al., 2006; Sun and Xu, 2009). These genomic differences can cause frequent chromosome nondisjunction during meiosis within the basidium during sexual reproduction, resulting in aneuploid progeny (Sun and Xu, 2009). Furthermore, meiotic recombination occurs at a much lower rate during inter-variety mating compared to crosses between serotype D strains (Sun and Xu, 2007). In addition, mitotic recombination was found to occur in the course of replicating the daughter nuclei within the basidium. Among 230 AD hybrids generated via a laboratory-defined mating, Vogan et al. found a large number of recombinant genotypes that could not be explained by the four haplotypes generated by one round of meiosis

(Vogan et al., 2013). This led to the discovery of mitotic recombination in the form of chromosome loss and crossing over as a novel mechanism that generates recombinant genotypes during inter-varietal crosses. Similarly, the sequenced genomes of CGSC and CNSC show 15-17% sequence divergence with multiple chromosomal rearrangements (D'Souza et al., 2011). Whether such rearrangements contribute to reduced recombination and chromosomal nondisjunction in CGSC x CNSC crosses remains to be investigated.

Genomes of AD hybrids are highly variable with evidence of homozygosity (or hemizygosity) interspersed with heterozygosity across the genome (Sun and Xu, 2009; Vogan et al., 2016, 2013). Laboratory investigations have shown that homozygosity could be derived through either chromosome loss or mitotic gene conversion leading to loss of heterozygosity (Hu et al., 2008; Li et al., 2012). Interestingly, natural AD hybrids show preferential retention of specific alleles and chromosomes from one of the two parents, suggesting that those alleles may offer survival and growth benefits under specific conditions (Hu et al., 2008). Even when both mating types are present in the genome, cryptococcal hybrids are rarely self-fertile. Basidiospores produced by three self-fertile **a**ADα hybrids germinated at a very low rate of ~5% in laboratory conditions: these three AD hybrids did not produce any sexual spores when co-incubated with haploid *MAT***a** and *MAT*α strains belonging to serotypes A and D (Lengeler et al., 2001). While hybrid infertility may represent an evolutionary dead-end for hybridization in plants and animals, this is not the case for *Cryptococcus* sp. and other fungi where sexually infertile hybrids can still reproduce asexually through mitosis and budding.

5.10 Some cryptococcal hybrids display transgressive segregation, including hybrid vigor

In genetics, the formation of extreme phenotypes, observed in segregated hybrid populations compared to phenotypes observed in the parental lines, is called transgressive segregation. These extreme (transgressive) phenotypes can be either positive or negative in terms of fitness in comparison to parental populations. Typically, hybrid populations will show more variation in both genotype and gene expression levels than either parental population. As a result, hybrid populations will

have some traits that are extreme (transgressive) in nature. If both parents' favorable alleles come together, the hybrids may have higher fitness than the two parents. In addition, the genetic diversity and associated phenotypic variation among hybrids may allow the hybrids to populate different environments/niches in which the parental populations do not reside, or directly compete with parental populations in the existing environments.

In CNSC and CGSC, there are three major virulence factors: the ability to grow at or above 37ºC which enables growth within mammalian hosts, production of a protective polysaccharide capsule around the cells, and production of melanin pigments with antioxidant properties (Charlier et al., 2005; Doering et al., 1999; Fromtling et al., 1982; Kwon-Chung and Rhodes, 1986; McGill et al., 2009; Perfect et al., 2006; Salas et al., 1996; Venn-Watson et al., 2012). Studies on virulence and fitness of cryptococcal hybrids have obtained variable results with some studies pointing to hybrid vigor while others finding no difference or even reduced virulence compared to parents. Chaturvedi et al. found 8 environmental and clinical AD hybrids to be similar in virulence to H99, a highly virulent serotype A *MAT*α strain, in mouse models (Chaturvedi et al., 2002). In comparison, the isogenic serotype D pair, JEC20 (D**a**) and JEC21 (Dα), did not cause death in these mice. Similarly, a laboratory-constructed αADα hybrid was found to be comparable in virulence to the H99 parent whereas the other parent, JEC21, did not cause any deaths in the mouse inhalation model (Lin et al., 2007). Another study found hybrid vigor in six naturally occurring and three laboratory constructed **a**ADα strains where they were significantly more resistant to UV irradiation than H99, JEC21 and serotype A**a** strains from Botswana (Litvintseva et al., 2007). Similarly, the majority of 31 global natural AD isolates tested were found to be resistant to the antifungal drug FK506 (W. Li et al., 2012).

In a quantitative phenotypic analysis, Vogan et al. generated 230 AD hybrids from a laboratory-defined cross between JEC20 (D**a**) and CDC15 (Aα) and assayed their melanin production, capsule production, cell size and susceptibility to the antifungal fluconazole (Vogan et al., 2016). JEC20 and CDC15 significantly differed from each other in these four traits. The majority of hybrid offspring produced melanin levels that were intermediate between the two parents whereas several progeny strains

produced significantly more or less melanin than the parents. A similar pattern was observed for fluconazole resistance. In contrast, most hybrid progeny were significantly larger in cell size than their haploid parents. Capsule production was observed to be highly heterogeneous for both parental and hybrid progeny strains with only a proportion of cells within each strain producing capsule. However, the proportion of capsule-producing cells in each sample as examined under the microscope was very consistent across repeated testing. Thus, instead of measuring the capsule size of individual cells, the proportion of cells of each strain that produced any amount of capsule was calculated. Under the tested conditions, JEC20 did not produce any capsule while CDC15 produced some: similarly, most AD hybrids produced very little to no capsule. However, a small number of progeny strains exhibited significantly higher capsule production with the highest proportion reaching 60%. Together, these results suggest that some cryptococcal hybrids can show hybrid vigor and transgressive segregation at least *in vitro* and in the *in vivo* murine models.

Compared to the systematic laboratory investigations, fitness of cryptococcal hybrids in clinical settings has not been studied as frequently. In one clinical observational study conducted in France, 483 cryptococcal isolates were recovered from 234 cryptococcosis patients between 1997 and 2001, out of which 20% were determined to be AD hybrids (Desnos-Ollivier et al., 2015). Clinical presentation and disease severity were statistically similar between infections caused by serotypes A, D and AD strains even though there was a trend towards AD hybrids causing less disseminated infection and less frequent abnormal lung imaging in patients. Patients infected with AD strains were significantly more likely to achieve CSF sterilization within two weeks of starting antifungal therapy compared to serotypes A and D. This study suggested that in human hosts in France, AD hybrids cause less severe infections than haploid serotype A and D strains.

The hybrid vigor displayed by some cryptococcal hybrids may be attributed at least partly to their higher ploidy compared to the haploid parents. Using progeny derived from unisexual α-α mating of a hypersexual Dα strain, it was shown that wild-type phenotypes were restored when aneuploid progeny strains became haploid due to loss of chromosomes (Ni et al., 2013). In the aneuploid state, these six progeny strains

were resistant to the antifungal fluconazole and outgrew the parent strain in a competitive growth assay by a ratio of ~4:1 (Ni et al., 2013). The researchers showed that aneuploidy was responsible for the enhanced expression of the virulence factors and that loss of heterozygosity reinstated wild-type phenotypes. A study of 19 AD hybrids by Cogliati et al. found virulence in mice to be positively correlated with the level of heterozygosity of the strains (Cogliati et al., 2011). The *MAT* locus has also been associated with increased virulence of cryptococcal hybrids. Using laboratory-constructed AD hybrids, it was shown that hybrids heterozygous at the *MAT* locus and homozygous for *MATα* showed no change in virulence potential whereas hybrids homozygous for ***MATa*** had significantly reduced virulence (Lin et al., 2008). Other findings also suggest that the presence of *MATα* allele from serotype A parent in AD hybrids is correlated with increased virulence whereas the presence of the ***MATa*** allele from either serotype A or serotype D parent is correlated with moderate to no virulence (Barchiesi et al., 2005).

The diversity of observations on virulence potential of cryptococcal hybrids suggest that their fitness and degree of virulence can vary significantly between individual hybrids. Such variations are understandable. Hybridization between divergent cryptococcal lineages with large genetic differences can produce a diversity of novel allelic combinations that could impact the expression of virulence traits and other phenotypes, generating frequent transgressive segregations (Shahid et al. 2008).

5.11 *C. neoformans* and *C. gattii* species complexes as a model system for hybridization

As shown in the preceding sections, the sister species complexes CNSC and CGSC present an excellent model system for the study of hybridization. Many well-characterized tools, strains, markers and assays are available in this model system. For example, the sexual cycles of CNSC and CGSC have been fully elucidated and many strains can readily mate with each other in the laboratory. The spore viability of laboratory-defined mating between strains of serotypes A and D typically ranges from 5%-30% (Forsythe et al., 2016; Lengeler et al., 2001; Vogan et al., 2013). Viable sexual spores from mating between CGSC and CNSC can be obtained despite lower germination rates (Chaturvedi et al., 2002). Hundreds of fully-sequenced cryptococcal

genomes encompassing all lineages are available on NCBI (for example Rhodes et al., 2017) making genetic analyses readily accessible. Genetic manipulations of cryptococcal strains can be achieved using well-characterized gene knockout systems (Arras et al., 2015; Lin et al., 2015) and more recently developed Crispr-Cas9 system for targeted gene editing in CNSC (Fan and Lin, 2018). The virulence traits of cryptococcal strains can be quantified using standardized assays for melanin production (Hopfer and Blank, 1976; Vogan et al., 2016), capsule production (Zaragoza et al., 2008), antifungal resistance (Rex et al., 2008) and high temperature growth (Perfect et al., 2006). The use of murine models of cryptococcosis has been optimized to study the virulence of cryptococcal strains *in vivo* (Sabiiti et al., 2012; Thompson et al., 2012). In addition, waxworms (*Galleria mellonella*) were also shown to be a good and cost-effective *in vivo* model for virulence testing (Fuchs et al., 2010).

The genetically diverged yet still closely related lineages of CNSC and CGSC are representative of other taxa that have and/or are currently undergoing hybridization in the wild. The added benefit of short generation times makes it convenient to generate and follow the evolutionary trajectories of cryptococcal hybrids in the laboratory (Sun and Xu, 2007; Vogan et al., 2013, 2016). As described above, the findings on cryptococcal hybrids have shed light on not only the potential consequences of hybridization in a clinical setting but also its role in shaping the evolution of this fungal pathogen.

5.12 Concluding remarks

Hybridization between closely related but evolutionary divergent organisms can often have unpredictable outcomes and significantly affect the evolutionary trajectory of the involved lineages. The increasing prevalence of AD hybrids in natural environments and in patients suggests that these hybrids will play increasingly important roles in natural cryptococcal populations. Unlike in plants and animals, aneuploidy and self-infertility are not impediments to the survival of cryptococcal hybrids as they can successfully propagate through mitosis and asexual budding. In fact, the novelty and plasticity of their genomes can give some hybrids a competitive advantage over parental strains. Given that AD hybrids are responsible for a substantial amount of

cryptococcal infections, it is important to understand how the unique genetic makeup of hybrids translates to increased fitness in the environment.

Our goal in writing this review was to provide a comprehensive yet concise overview of the current body of knowledge on hybridization and hybrids of the *C. neoformans* and *C. gattii* species complexes. We hope that this will stimulate more investigations into less well understood aspects of these hybrids. For example, evidence for both historical and recent hybridization events have been found (D'Souza et al., 2011; Sharpton et al., 2008; Xu et al., 2000; 2002): it would be interesting to identify the prevalence and timeline of such hybridization events and their impacts on the genome architectures throughout the history of CNSC and CGSC. The availability of hundreds of sequenced genomes from diverse cryptococcal lineages should make this feasible. Similarly, virulence of hybrids is highly variable both in animal models and clinical observations of human patients. At present, the genetic bases for such variations are little understood. The increasing prevalence of AD hybrids seems to suggest that at least some hybrids have an adaptive advantage over their parental lineages. While the inconsistency of these findings can be explained by differences in genetic makeup between hybrid strains, further investigations will be able to shed light on the specific genetic factors and genotype-environment interactions responsible for the increased fitness of the hybrids. Finally, as demonstrated by Vogan et al. (2016), hybrids can be successfully utilized to identify genetic factors that contribute to differential expression of virulence factors in haploid serotype A and D parent strains. Therefore, cryptococcal hybrids are not only an interesting study in themselves, they can be used as a tool to gain insights into haploid lineages of CNSC and CGSC as well as hybridization in general in other taxa.

5.13 References

Alvarez, M., Casadevall, A., 2007. Cell-to-cell spread and massive vacuole formation after *Cryptococcus neoformans* infection of murine macrophages. BMC Immunol. 8, 16. https://doi.org/10.1186/1471-2172-8-16

Aminnejad, M., Diaz, M., Arabatzis, M. et al. 2012. Identification of Novel Hybrids Between *Cryptococcus neoformans* var. grubii VNI and *Cryptococcus gattii* VGII. Mycopathologia 173, 337–346. https://doi.org/10.1007/s11046-011-9491-x

Arras, S.D.M., Chitty, J.L., Blake, K.L. et al. 2015. A Genomic Safe Haven for Mutant Complementation in *Cryptococcus neoformans*. PLoS One 10, e0122916. https://doi.org/10.1371/journal.pone.0122916

Arsenijevic, V.A., Pekmezovic, M.G., Meis, J.F., Hagen, F., 2014. Molecular epidemiology and antifungal susceptibility of Serbian *Cryptococcus neoformans* isolates. Mycoses 57, n/a-n/a. https://doi.org/10.1111/myc.12171

Barchiesi, F., Cogliati, M., Esposto, M.C. et al. 2005. Comparative analysis of pathogenicity of *Cryptococcus neoformans* serotypes A, D and AD in murine cryptococcosis. J. Infect. 51, 10–16. https://doi.org/10.1016/j.jinf.2004.07.013

Baroni, F.D.A., Paula, C.R., Da Silva, É.G. et al. 2006. *Cryptococcus neoformans* strains isolated from church towers in Rio de Janeiro City, RJ, Brazil. Rev. Inst. Med. Trop. Sao Paulo 48, 71–75. https://doi.org//S0036-46652006000200003

Barreto de Oliveira, M.T., Boekhout, T., Theelen, B. et al. 2004. *Cryptococcus neoformans* shows a remarkable genotypic diversity in Brazil. J. Clin. Microbiol. 42, 1356–9. https://doi.org/10.1128/JCM.42.3.1356-1359.2004

Bennett, J.E., Kwon-Chung, K.J., Howard, D.H., 1977. Epidemiologic differences among serotypes of *Cryptococcus neoformans*. Am. J. Epidemiol. 105, 582–6.

Bhattacharjee, a K., Bennett, J.E., Glaudemans, C.P., 1984. Capsular polysaccharides of *Cryptococcus neoformans*. Rev. Infect. Dis. 6, 619–24.

Boekhout, T., Theelen, B., Diaz, M. et al. 2001. Hybrid genotypes in the pathogenic

yeast *Cryptococcus neoformans*. Microbiology 147, 891–907. https://doi.org/10.1099/00221287-147-4-891

Bovers, M., Hagen, F., Kuramae, E.E., Diaz, M.R., Spanjaard, L., Dromer, F., Hoogveld, H.L., Boekhout, T., 2006. Unique hybrids between the fungal pathogens *Cryptococcus neoformans* and *Cryptococcus gattii*. FEMS Yeast Res. 6, 599–607. https://doi.org/10.1111/j.1567-1364.2006.00082.x

Bovers, M., Hagen, F., Kuramae, E.E., Hoogveld, H.L., Dromer, F., St-Germain, G., Boekhout, T., 2008. AIDS patient death caused by novel *Cryptococcus neoformans* x C. gattii hybrid. Emerg. Infect. Dis. 14, 1105–8. https://doi.org/10.3201/eid1407.080122

Busse, O., 1894. Uber parasitare zelleinschlusse und ihre zuchtung. Zentralbl Bakteriol 16, 175–180.

Campbell, L.T., Fraser, J.A., Nichols, C.B. et al. 2005. Clinical and environmental isolates of *Cryptococcus gattii* from Australia that retain sexual fecundity. Eukaryot. Cell 4, 1410–9. https://doi.org/10.1128/EC.4.8.1410-1419.2005

Casadevall, A., Freij, J.B., Hann-Soden, C., Taylor, J., 2017. Continental Drift and Speciation of the *Cryptococcus neoformans* and *Cryptococcus gattii* Species Complexes. mSphere 2. https://doi.org/10.1128/mSphere.00103-17

Charlier, C., Chrétien, F., Baudrimont, M. et al. 2005. Capsule structure changes associated with *Cryptococcus neoformans* crossing of the blood-brain barrier. Am. J. Pathol. 166, 421–432. https://doi.org/10.1016/S0002-9440(10)62265-1

Charlier, C., Nielsen, K., Daou, S., Brigitte, M., Chretien, F., Dromer, F., 2009. Evidence of a role for monocytes in dissemination and brain invasion by *Cryptococcus neoformans*. Infect. Immun. 77, 120–127. https://doi.org/10.1128/IAI.01065-08

Chaturvedi, V., Fan, J., Stein, B., Behr, M.J., Samsonoff, W.A., Wickes, B.L., Chaturvedi, S., 2002. Molecular genetic analyses of mating pheromones reveal intervariety mating or hybridization in *Cryptococcus neoformans*. Infect. Immun. 70, 5225–35. https://doi.org/10.1128/IAI.70.9.5225-5235.2002

Chau, T.T., Mai, N.H., Phu, N.H. et al. 2010. A prospective descriptive study of cryptococcal meningitis in HIV uninfected patients in Vietnam - high prevalence of *Cryptococcus neoformans* var grubii in the absence of underlying disease. BMC Infect. Dis. 10, 199. https://doi.org/10.1186/1471-2334-10-199

Chen, J., Varma, A., Diaz, M.R., Litvintseva, A.P., Wollenberg, K.K., Kwon-Chung, K.J., 2008. *Cryptococcus neoformans* strains and infection in apparently immunocompetent patients, China. Emerg. Infect. Dis. 14, 755–762. https://doi.org/10.3201/eid1405.071312

Chen, M., Li, X.R., Wu, S.X. et al. 2010. Molecular epidemiology of *Cryptococcus neoformans* species complex isolates from HIV-positive and HIV-negative patients in southeast China. Front. Med. China 4, 117–126. https://doi.org/10.1007/s11684-010-0011-z

Chen, S.C.-A., Meyer, W., Sorrell, T.C., 2014. *Cryptococcus gattii* infections. Clin. Microbiol. Rev. 27, 980–1024. https://doi.org/10.1128/CMR.00126-13

Chen, S.C.A., Slavin, M.A., Heath, C.H. et al. 2012. Clinical Manifestations of *Cryptococcus gattii* Infection: Determinants of Neurological Sequelae and Death. Clin. Infect. Dis. 55, 789–798. https://doi.org/10.1093/cid/cis529

Choi, Y.H., Ngamskulrungroj, P., Varma, A. et al. 2010. Prevalence of the VNIc genotype of *Cryptococcus neoformans* in non-HIV-associated cryptococcosis in the Republic of Korea. FEMS Yeast Res. 10, 769–778. https://doi.org/10.1111/j.1567-1364.2010.00648.x

Chowdhary, A., Randhawa, H.S., Prakash, A., Meis, J.F., 2012. Environmental prevalence of *Cryptococcus neoformans* and *Cryptococcus gattii* in India: An update. Crit. Rev. Microbiol. 38, 1–16. https://doi.org/10.3109/1040841X.2011.606426

Chowdhary, A., Randhawa, H.S., Sundar, G. et al. 2011. In vitro antifungal susceptibility profiles and genotypes of 308 clinical and environmental isolates of *Cryptococcus neoformans* var. grubii and *Cryptococcus gattii* serotype B from north-western India. J. Med. Microbiol. 60, 961–967.

https://doi.org/10.1099/jmm.0.029025-0

Christman, M.C., Kedwaii, A., Xu, J., Donis, R.O., Lu, G., 2011. Pandemic (H1N1) 2009 virus revisited: an evolutionary retrospective. Infect. Genet. Evol. 11, 803–11. https://doi.org/10.1016/j.meegid.2011.02.021

Cogliati, M., 2013. Global Molecular Epidemiology of *Cryptococcus neoformans* and *Cryptococcus gattii*: An Atlas of the Molecular Types. Scientifica (Cairo). https://doi.org/10.1155/2013/675213

Cogliati, M., Barchiesi, F., Spreghini, E., Tortorano, A.M., 2011. Heterozygosis and Pathogenicity of *Cryptococcus neoformans* AD-Hybrid Isolates. Mycopathologia 173, 347–357. https://doi.org/10.1007/s11046-011-9467-x

Cogliati, M., D'Amicis, R., Zani, A. et al. 2016. Environmental distribution of *Cryptococcus neoformans* and C. Gattii around the Mediterranean basin. FEMS Yeast Res. 16, 1–12. https://doi.org/10.1093/femsyr/fow045

Cogliati, M., Esposto, M.C., Clarke, D.L., Wickes, B.L., Viviani, M.A., 2001. Origin of *Cryptococcus neoformans* var. *neoformans* diploid strains. J. Clin. Microbiol. 39, 3889–94. https://doi.org/10.1128/JCM.39.11.3889-3894.2001

Cogliati, M., Esposto, M.C., Tortorano, A.M., Viviani, M.A., 2006. *Cryptococcus neoformans* population includes hybrid strains homozygous atmating-type locus. FEMS Yeast Res. 6, 608–613. https://doi.org/10.1111/j.1567-1364.2006.00085.x

D'Souza, C.A., Kronstad, J.W., Taylor, G. et al. 2011. Genome variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts. MBio 2, e00342-10. https://doi.org/10.1128/mBio.00342-10

Del Valle, L., Pina-Oviedo, S., 2006. HIV disorders of the brain; pathology and pathogenesis. Front. Biosci. 11, 718–732.

Desnos-Ollivier, M., Patel, S., Raoux-Barbot, D. et al. 2015. Cryptococcosis Serotypes Impact Outcome and Provide Evidence of *Cryptococcus neoformans* Speciation. MBio 6, e00311. https://doi.org/10.1128/mBio.00311-15

Doering, T.L., Nosanchuk, J.D., Roberts, W.K., Casadevall, A., 1999. Melanin as a

potential cryptococcal defence against microbicidal proteins. Med. Mycol. 37, 175–181. https://doi.org/10.1046/j.1365-280X.1999.00218.x

Ellabib, M.S., Aboshkiwa, M.A., Husien, W.M. et al. 2016. Isolation, Identification and Molecular Typing of *Cryptococcus neoformans* from Pigeon Droppings and Other Environmental Sources in Tripoli, Libya. Mycopathologia 181, 603–608. https://doi.org/10.1007/s11046-016-9996-4

Escandon, P., Sanchez, A., Marti-nez, M. et al. 2006. Molecular epidemiology of clinical and environmental isolates of the *Cryptococcus neoformans* species complex reveals a high genetic diversity and the presence of the molecular type VGII mating type a in Colombia. FEMS Yeast Res. 6, 625–635. https://doi.org/10.1111/j.1567-1364.2006.00055.x

Fan, Y., Lin, X., 2018. Multiple Applications of a Transient CRISPR-Cas9 Coupled with Electroporation (TRACE) System in the *Cryptococcus neoformans* Species Complex. Genetics 208, 1357–1372. https://doi.org/10.1534/genetics.117.300656

Favalessa, O.C., de Paula, D.A.J., Dutra, V. et al. 2014. Molecular typing and in vitro antifungal susceptibility of *Cryptococcus* spp from patients in Midwest Brazil. J. Infect. Dev. Ctries. https://doi.org/10.3855/jidc.4446

Feldmesser, M., Kress, Y., Novikoff, P., Casadevall, A., 2000. *Cryptococcus neoformans* is a facultative intracellular pathogen in murine pulmonary infection. Infect. Immun. 68, 4225–4237. https://doi.org/10.1128/IAI.68.7.4225-4237.2000

Feng, X., Yao, Z., Ren, D., Liao, W., Wu, J., 2008. Genotype and mating type analysis of *Cryptococcus neoformans* and *Cryptococcus gattii* isolates from China that mainly originated from non-HIV-infected patients. FEMS Yeast Res. 8, 930–938. https://doi.org/10.1111/j.1567-1364.2008.00422.x

Ferreira, A.S., Sampaio, A., Maduro, A.P., Silva, I., Teles, F., Martins, M. da L., Inácio, J., 2014. Genotypic diversity of environmental *Cryptococcus neoformans* isolates from Northern Portugal. Mycoses 57, 98–104. https://doi.org/10.1111/myc.12106

Forman, H.J., Torres, M., 2002. Reactive oxygen species and cell signaling:

Respiratory burst in macrophage signaling, in: American Journal of Respiratory and Critical Care Medicine. https://doi.org/10.1164/rccm.2206007

Forsythe, A., Vogan, A., Xu, J., 2016. Genetic and environmental influences on the germination of basidiospores in the *Cryptococcus neoformans* species complex. Sci. Rep. 6, 33828. https://doi.org/10.1038/srep33828

Franzot, S.P., Salkin, I.F., Casadevall, A., 1999. *Cryptococcus neoformans* var. *grubii*: separate varietal status for *Cryptococcus neoformans* serotype A isolates. J. Clin. Microbiol. 37, 838–40.

Fraser, J. a, Giles, S.S., Wenink, E.C., Geunes-Boyer, S.G., Wright, J.R., Diezmann, S., Allen, A., Stajich, J.E., Dietrich, F.S., Perfect, J.R., Heitman, J., 2005. Same-sex mating and the origin of the Vancouver Island *Cryptococcus gattii* outbreak. Nature 437, 1360–1364. https://doi.org/10.1038/nature04220

Fraser, J.A., Subaran, R.L., Nichols, C.B., Heitman, J., 2003. Recapitulation of the sexual cycle of the primary fungal pathogen *Cryptococcus neoformans* var. *gattii*: implications for an outbreak on Vancouver Island, Canada. Eukaryot. Cell 2, 1036–45. https://doi.org/10.1128/EC.2.5.1036-1045.2003

Freire, A.K.L., dos Santos Bentes, A., de Lima Sampaio, I. et al. 2012. Molecular characterisation of the causative agents of Cryptococcosis in patients of a tertiary healthcare facility in the state of Amazonas-Brazil. Mycoses 55, e145–e150. https://doi.org/10.1111/j.1439-0507.2012.02173.x

Fromtling, R.A., Shadomy, H.J., Jacobson, E.S., 1982. Decreased virulence in stable, acapsular mutants of *Cryptococcus neoformans*. Mycopathologia 79, 23–29. https://doi.org/10.1007/BF00636177

Fuchs, B.B., O'Brien, E., Khoury, J.B. El, Mylonakis, E., 2010. Methods for using *Galleria mellonella* as a model host to study fungal pathogenesis. Virulence 1, 475–482. https://doi.org/10.4161/viru.1.6.12985

Gago, S., Serrano, C., Alastruey-Izquierdo, A. et al. 2017. Molecular identification, antifungal resistance and virulence of *Cryptococcus neoformans* and *Cryptococcus deneoformans* isolated in Seville, Spain. Mycoses 60, 40–50.

https://doi.org/10.1111/myc.12543

Goldman, D., Lee, S.C., Casadevall, A., 1994. Pathogenesis of pulmonary *Cryptococcus neoformans* infection in the rat. Infect. Immun. 62, 4755–4761.

Goulet, B.E., Roda, F., Hopkins, R., 2017. Hybridization in Plants: Old Ideas, New Techniques. Plant Physiol. 173, 65–78. https://doi.org/10.1104/pp.16.01340

Hagen, F., Hare Jensen, R., Meis, J.F. et al. 2016. Molecular epidemiology and in vitro antifungal susceptibility testing of 108 clinical *Cryptococcus neoformans* sensu lato and *Cryptococcus gattii* sensu lato isolates from Denmark. Mycoses 59, 576–584. https://doi.org/10.1111/myc.12507

Hagen, F., Illnait-Zaragozí, M.-T., Meis, J.F. et al. 2012. Extensive genetic diversity within the Dutch clinical *Cryptococcus neoformans* population. J. Clin. Microbiol. 50, 1918–26. https://doi.org/10.1128/JCM.06750-11

Hagen, F., Khayhan, K., Theelen, B., Kolecka, A., Polacheck, I., Sionov, E., Falk, R., Parnmen, S., Lumbsch, H.T., Boekhout, T., 2015. Recognition of seven species in the *Cryptococcus gattii*/*Cryptococcus neoformans* species complex. Fungal Genet. Biol. 78, 16–48. https://doi.org/10.1016/j.fgb.2015.02.009

Hagen, F., Lumbsch, H.T., Arsic Arsenijevic, V. et al. 2017. Importance of Resolving Fungal Nomenclature: the Case of Multiple Pathogenic Species in the *Cryptococcus* Genus. mSphere 2. https://doi.org/10.1128/mSphere.00238-17

Hiremath, S.S., Chowdhary, A., Kowshik, T., Randhawa, H.S., Sun, S., Xu, J., 2008. Long-distance dispersal and recombination in environmental populations of *Cryptococcus neoformans* var. *grubii* from India. Microbiology 154, 1513–1524. https://doi.org/10.1099/mic.0.2007/015594-0

Hironaga, M., Ikeda, R., Fukazawa, Y., Watanabe, S., 1983. Mating types and serotypes of *Cryptococcus neoformans* isolated in Japan. Sabouraudia 21, 73–8.

Hopfer, R.L., Blank, F., 1976. Caffeic acid-containing medium for identification of *Cryptococcus neoformans*. J. Clin. Microbiol. 2, 115–20.

Hu, G., Liu, I., Sham, A., Stajich, J.E., Dietrich, F.S., Kronstad, J.W., 2008.

Comparative hybridization reveals extensive genome variation in the AIDS-associated pathogen *Cryptococcus neoformans*. Genome Biol. 9, R41. https://doi.org/10.1186/gb-2008-9-2-r41

Husain, S., Wagener, M.M., Singh, N., 2001. *Cryptococcus neoformans* infection in organ transplant recipients: variables influencing clinical characteristics and outcome. Emerg. Infect. Dis. https://doi.org/10.3201/eid0703.010302

Idnurm, A., 2010. A tetrad analysis of the basidiomycete fungus *Cryptococcus neoformans*. Genetics 185, 153–63. https://doi.org/10.1534/genetics.109.113027

Kaocharoen, S., Ngamskulrungroj, P., Firacative, C. et al. 2013. Molecular epidemiology reveals genetic diversity amongst isolates of the *Cryptococcus neoformans*/*C. gattii* species complex in Thailand. PLoS Negl. Trop. Dis. 7, e2297. https://doi.org/10.1371/journal.pntd.0002297

Kassi, F.K., Drakulovski, P., Bellet, V. et al. 2016. Molecular epidemiology reveals genetic diversity among 363 isolates of the *Cryptococcus neoformans* and *Cryptococcus gattii* species complex in 61 Ivorian HIV-positive patients. Mycoses 59, 811–817. https://doi.org/10.1111/myc.12539

Kavanaugh, L.A., Fraser, J.A., Dietrich, F.S., 2006. Recent evolution of the human pathogen *Cryptococcus neoformans* by intervarietal transfer of a 14-Gene fragment. Mol. Biol. Evol. 23, 1879–1890. https://doi.org/10.1093/molbev/msl070

Keller, S.M., Viviani, M.A., Esposto, M.C., Cogliati, M., Wickes, B.L., 2003. Molecular and genetic characterization of a serotype A MATa*Cryptococcus neoformans* isolate. Microbiology 149, 131–142. https://doi.org/10.1099/mic.0.25921-0

Kidd, S.E., Hagen, F., Tscharke, R.L. et al. 2004. A rare genotype of *Cryptococcus gattii* caused the cryptococcosis outbreak on Vancouver Island (British Columbia, Canada). Proc. Natl. Acad. Sci. U. S. A. 101, 17258–17263. https://doi.org/10.1073/pnas.0402981101

Krogerus, K., Magalhães, F., Vidgren, V., Gibson, B., 2017. Novel brewing yeast

hybrids: creation and application. Appl. Microbiol. Biotechnol. 101, 65–78. https://doi.org/10.1007/s00253-016-8007-5

Kwon-Chung, K.J., 1976a. Morphogenesis of *Filobasidiella neoformans*, the sexual state of *Cryptococcus neoformans*. Mycologia 68, 821. https://doi.org/10.2307/3758800

Kwon-Chung, K.J., 1976b. A new species of *Filobasidiella*, the sexual state of *Cryptococcus neoformans* B and C serotypes. Mycologia 68, 943–6.

Kwon-Chung, K.J., 1975. A new genus, *Filobasidiella*, the perfect state of *Cryptococcus neoformans*. Mycologia 67, 1197. https://doi.org/10.2307/3758842

Kwon-Chung, K.J., Bennett, J.E., 1984. High prevalence of *Cryptococcus neoformans* var. *gattii* in tropical and subtropical regions. Zentralblatt für Bakteriol. Mikrobiol. und Hyg. 257, 213–218.

Kwon-Chung, K.J., Bennett, J.E., Wickes, B.L. et al. 2017. The Case for Adopting the "Species Complex" Nomenclature for the Etiologic Agents of Cryptococcosis. mSphere 2:e00357-16. https://doi.org/10.1128/mSphere.00357-16.

Kwon-chung, K.J., Boekhout, T., Fell, J.W., Diaz, M., 2002. Proposal to conserve the name *Cryptococcus gattii* against *C* . *hondurianus* and *C* . *bacillisporus* (Basidiomycota , Tremellomycetidae). Taxon 51, 804–806.

Kwon-Chung, K.J., Rhodes, J.C., 1986. Encapsulation and melanin formation as indicators of virulence in *Cryptococccus neoformans*. Infect. Immun. 51, 218–223.

Kwon-Chung, K.J., Varma, A., 2006. Do major species concepts support one, two or more species within *Cryptococcus neoformans*? FEMS Yeast Res. 6, 574–587. https://doi.org/10.1111/j.1567-1364.2006.00088.x

Lengeler, K.B., Cox, G.M., Heitman, J., 2001. Serotype AD strains of *Cryptococcus neoformans* are diploid or aneuploid and are heterozygous at the mating-type locus. Infect. Immun. 69, 115–22. https://doi.org/10.1128/IAI.69.1.115-122.2001

Lengeler, K.B., Fox, D.S., Fraser, J.A. et al. 2002. Mating-type locus of *Cryptococcus*

neoformans: a step in the evolution of sex chromosomes. Eukaryot. Cell 1, 704–18. https://doi.org/10.1128/EC.1.5.704-718.2002

Lengeler, K.B., Wang, P., Cox, G.M., Perfect, J.R., Heitman, J., 2000. Identification of the *MAT**a*** mating-type locus of *Cryptococcus neoformans* reveals a serotype A *MAT**a*** strain thought to have been extinct.

Li, M., Liao, Y., Chen, M., Pan, W., Weng, L., 2012. Antifungal susceptibilities of *Cryptococcus* species complex isolates from AIDS and non-AIDS patients in Southeast China. Brazilian J. Infect. Dis. https://doi.org/10.1016/S1413-8670(12)70301-X

Li, W., Averette, A.F., Desnos-Ollivier, M., Ni, M., Dromer, F., Heitman, J., 2012. Genetic Diversity and Genomic Plasticity of *Cryptococcus neoformans* AD Hybrid Strains. G3 (Bethesda). 2, 83–97. https://doi.org/10.1534/g3.111.001255

Libkind, D., Hittinger, C.T., Valerio, E., Goncalves, C., Dover, J., Johnston, M., Goncalves, P., Sampaio, J.P., 2011. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. Proc. Natl. Acad. Sci. https://doi.org/10.1073/pnas.1105430108

Lin, X., Chacko, N., Wang, L., Pavuluri, Y., 2015. Generation of stable mutants and targeted gene deletion strains in *Cryptococcus neoformans* through electroporation. Med. Mycol. 53, 225–234. https://doi.org/10.1093/mmy/myu083

Lin, X., Hull, C.M., Heitman, J., 2005. Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*. Nature 434, 1017–1021. https://doi.org/10.1038/nature03448

Lin, X., Litvintseva, A.P., Nielsen, K., Patel, S., Floyd, A., Mitchell, T.G., Heitman, J., 2007. αADα Hybrids of *Cryptococcus neoformans*: Evidence of Same-Sex Mating in Nature and Hybrid Fitness. PLoS Genet. 3, e186. https://doi.org/10.1371/journal.pgen.0030186

Lin, X., Nielsen, K., Patel, S., Heitman, J., 2008. Impact of mating type, serotype, and ploidy on the virulence of *Cryptococcus neoformans*. Infect. Immun. 76, 2923–38. https://doi.org/10.1128/IAI.00168-08

Lin, X., Patel, S., Litvintseva, A.P., Floyd, A., Mitchell, T.G., Heitman, J., 2009. Diploids in the *Cryptococcus neoformans* Serotype A Population Homozygous for the α Mating Type Originate via Unisexual Mating. PLoS Pathog. 5, e1000283. https://doi.org/10.1371/journal.ppat.1000283

Litvintseva, A.P., Carbone, I., Rossouw, J., Thakur, R., Govender, N.P., Mitchell, T.G., 2011. Evidence that the Human Pathogenic Fungus *Cryptococcus neoformans* var. *grubii* May Have Evolved in Africa. PLoS One 6, e19688. https://doi.org/10.1371/journal.pone.0019688

Litvintseva, A.P., Kestenbaum, L., Vilgalys, R., Mitchell, T.G., 2005. Comparative Analysis of Environmental and Clinical Populations of *Cryptococcus neoformans*. J. Clin. Microbiol. 43, 556–564. https://doi.org/10.1128/JCM.43.2.556-564.2005

Litvintseva, A.P., Lin, X., Templeton, I., Heitman, J., Mitchell, T.G., 2007. Many globally isolated AD hybrid strains of *Cryptococcus neoformans* originated in Africa. PLoS Pathog. 3, 1109–1117. https://doi.org/10.1371/journal.ppat.0030114

Litvintseva, A.P., Marra, R.E., Nielsen, K., Heitman, J., Vilgalys, R., Mitchell, T.G., 2003. Evidence of sexual recombination among *Cryptococcus neoformans* serotype A isolates in sub-Saharan Africa. Eukaryot. Cell 2, 1162–8. https://doi.org/10.1128/EC.2.6.1162-1168.2003

Litvintseva, A.P., Thakur, R., Reller, L.B., Mitchell, T.G., 2005. Prevalence of Clinical Isolates of *Cryptococcus gattii* Serotype C among Patients with AIDS in Sub-Saharan Africa. J. Infect. Dis. 192, 888–892. https://doi.org/10.1086/432486

Litvintseva, A.P., Thakur, R., Vilgalys, R., Mitchell, T.G., 2006. Multilocus sequence typing reveals three genetic subpopulations of *Cryptococcus neoformans* var. *grubii* (serotype A), including a unique population in Botswana. Genetics 172, 2223–2238. https://doi.org/10.1534/genetics.105.046672

Lockhart, S.R., Iqbal, N., Harris, J.R. et al. 2013. *Cryptococcus gattii* in the United States: Genotypic Diversity of Human and Veterinary Isolates. PLoS One 8,

e74737. https://doi.org/10.1371/journal.pone.0074737

Ma, H., Croudace, J.E., Lammas, D. a, May, R.C., 2007. Direct cell-to-cell spread of a pathogenic yeast. BMC Immunol. 8, 15. https://doi.org/10.1186/1471-2172-8-15

Maduro, A.P., Mansinho, K., Teles, F., Silva, I., Meyer, W., Martins, M.L., Inácio, J., 2014. Insights on the Genotype Distribution Among *Cryptococcus neoformans* and *C. gattii* Portuguese Clinical Isolates. Curr. Microbiol. 68, 199–203. https://doi.org/10.1007/s00284-013-0452-0

McGill, S., Malik, R., Saul, N. et al. 2009. Cryptococcosis in domestic animals in Western Australia: a retrospective study from 1995-2006. Med. Mycol. 47, 625–39. https://doi.org/10.1080/13693780802512519

Meyer, W., Aanensen, D.M., Boekhout, T. et al. 2009. Consensus multi-locus sequence typing scheme for *Cryptococcus neoformans* and *Cryptococcus gattii*. Med. Mycol. https://doi.org/10.1080/13693780902953886

Meyer, W., Castañeda, A., Jackson, S., Huynh, M., Castañeda, E., IberoAmerican Cryptococcal Study Group, the I.C.S., 2003. Molecular typing of IberoAmerican *Cryptococcus neoformans* isolates. Emerg. Infect. Dis. 9, 189–95. https://doi.org/10.3201/eid0902.020246

Meyer, W., Marszewska, K., Amirmostofian, M. et al. 1999. Molecular typing of global isolates of *Cryptococcus neoformans* var. *neoformans* by polymerase chain reaction fingerprinting and randomly amplified polymorphic DNA - A pilot study to standardize techniques on which to base a detailed epidemiological survey, in: Electrophoresis. pp. 1790–1799. https://doi.org/10.1002/(SICI)1522-2683(19990101)20:8<1790::AID-ELPS1790>3.0.CO;2-2

Meyer, W., Mitchell, T.G., Freedman, E.Z., Vilgalys, R., 1993. Hybridization probes for conventional DNA fingerprinting used as single primers in the polymerase chain reaction to distinguish strains of *Cryptococcus neoformans*. J. Clin. Microbiol. 31, 2274–2280. https://doi.org/doi:0095-1137/93/092274-07$02.00/0

Miglia, K.J., Govender, N.P., Rossouw, J. et al. 2011. Analyses of pediatric isolates of *Cryptococcus neoformans* from South Africa. J. Clin. Microbiol. 49, 307–14.

https://doi.org/10.1128/JCM.01277-10

Mitchell, T.G., Castañeda, E., Nielsen, K., Wanke, B., Lazéra, M.S., 2011. Environmental Niches for *Cryptococcus neoformans* and *Cryptococcus gattii*, in: *Cryptococcus*. American Society of Microbiology, pp. 237–259. https://doi.org/10.1128/9781555816858.ch18

Mlinaric-Missoni, E., Hagen, F., Chew, W.H.M. et al. 2011. In vitro antifungal susceptibilities and molecular typing of sequentially isolated clinical *Cryptococcus neoformans* strains from Croatia. J. Med. Microbiol. 60, 1487–1495. https://doi.org/10.1099/jmm.0.031344-0

Montagna, M.T., De Donno, A., Caggiano, G. et al. 2018. Molecular characterization of *Cryptococcus neoformans* and *Cryptococcus gattii* from environmental sources and genetic comparison with clinical isolates in Apulia, Italy. Environ. Res. 160, 347–352. https://doi.org/10.1016/J.ENVRES.2017.09.032

Nathan, C., Shiloh, M.U., 2000. Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens. Pnas 97, 8841–8848. https://doi.org/10.1073/pnas.97.16.8841

Ni, M., Feretzaki, M., Li, W., Floyd-Averette, A., Mieczkowski, P., Dietrich, F.S., Heitman, J., 2013. Unisexual and Heterosexual Meiotic Reproduction Generate Aneuploidy and Phenotypic Diversity De Novo in the Yeast *Cryptococcus neoformans*. PLoS Biol. 11, e1001653. https://doi.org/10.1371/journal.pbio.1001653

Nishikawa, M.M., Lazera, M.S., Barbosa, G.G., Trilles, L., Balassiano, B.R., Macedo, R.C.L., Bezerra, C.C.F., Perez, M.A., Cardarelli, P., Wanke, B., 2003. Serotyping of 467 *Cryptococcus neoformans* isolates from clinical and environmental sources in Brazil: analysis of host and regional patterns. J. Clin. Microbiol. 41, 73–77.

Perfect, J.R. 2006. *Cryptococcus neoformans*: the yeast that likes it hot. FEMS Yeast Res. 6, 463–8. https://doi.org/10.1111/j.1567-1364.2006.00051.x

Phillips, P., Galanis, E., MacDougall, L. et al. 2015. Longitudinal Clinical Findings

and Outcome among *Cryptococcus gattii* Patients in British Columbia. Clin. Infect. Dis. 60, 1368–1376. https://doi.org/10.1093/cid/civ041

Pini, G., Faggi, E., Bravetti, E., 2017. Molecular Typing of Clinical and Environmental *Cryptococcus neoformans* Strains Isolated in Italy. Open J. Med. Microbiol. 07, 77–85. https://doi.org/10.4236/ojmm.2017.74007

Rajasingham, R., Smith, R.M., Park, B.J., Jarvis, J.N., Govender, N.P., Chiller, T.M., Denning, D.W., Loyse, A., Boulware, D.R., 2017. Global burden of disease of HIV-associated cryptococcal meningitis: an updated analysis. Lancet Infect. Dis. 17, 873–881. https://doi.org/10.1016/S1473-3099(17)30243-8

Rex, J.H., Alexander, B.D., Andes, D. et al. 2008. Reference method for broth dilution antifungal susceptibility testing of yeasts: approved standard - third edition. Clin. Lab. Stand. Inst. https://doi.org/10.1007/SpringerReference_5244

Rhodes, J., Desjardins, C.A., Sykes, S.M. et al. 2017. Population Genomics Of *Cryptococcus neoformans* var. grubii Reveals New Biogeographic Relationships And Finely Maps Hybridization. bioRxiv 132894. https://doi.org/10.1101/132894

Sabiiti, W., May, R.C., Pursall, E.R., 2012. Experimental models of cryptococcosis. Int. J. Microbiol. 2012, 626745. https://doi.org/10.1155/2012/626745

Salas, S.D., Bennett, J.E., Kwon-Chung, K.J., Perfect, J.R., Williamson, P.R., 1996. Effect of the laccase gene CNLAC1, on virulence of *Cryptococcus neoformans*. J. Exp. Med. 184, 377–386. https://doi.org/10.1084/jem.184.2.377

Sanchini, A., Smith, I.M., Sedlacek, L., Schwarz, R., Tintelnot, K., Rickerts, V., 2014. Molecular typing of clinical *Cryptococcus neoformans* isolates collected in Germany from 2004 to 2010. Med. Microbiol. Immunol. 203, 333–340. https://doi.org/10.1007/s00430-014-0341-6

Sanfelice, F., 1894. Contributo alla morfologia e biologia dei blastomiceti che si sviluppano nei succhi di alcuni frutti. Ann Ig. 4, 463–469.

Saracli, M.A., Yildiran, S.T., Sener, K. et al. 2006. Genotyping of Turkish

environmental *Cryptococcus neoformans* var. neoformans isolates by pulsed field gel electrophoresis and mating type. Mycoses 49, 124–129. https://doi.org/10.1111/j.1439-0507.2006.01203.x

Sharpton, T.J., Neafsey, D.E., Galagan, J.E., Taylor, J.W., 2008. Mechanisms of intron gain and loss in *Cryptococcus*. Genome Biol. 9, R24. https://doi.org/10.1186/gb-2008-9-1-r24

Smith, I.M., Stephan, C., Hogardt, M., Klawe, C., Tintelnot, K., Rickerts, V., 2015. Cryptococcosis due to *Cryptococcus gattii* in Germany from 2004–2013. Int. J. Med. Microbiol. 305, 719–723. https://doi.org/10.1016/J.IJMM.2015.08.023

Sorrell, T.C., 2001. *Cryptococcus neoformans* variety gattii. Med. Mycol. 39, 155–68. https://doi.org/10.1080/714031012

Steenbergen, J.N., Casadevall, A., 2000. Prevalence of *Cryptococcus neoformans* var. neoformans (Serotype D) and *Cryptococcus neoformans* var. grubii (Serotype A) isolates in New York City. J. Clin. Microbiol. 38, 1974–6.

Stie, J., Fox, D., 2012. Blood-brain barrier invasion by *Cryptococcus neoformans* is enhanced by functional interactions with plasmin. Microbiol. 158, 240–58. https://doi.org/10.1099/mic.0.051524-0

Sukroongreung, S., Nilakul, C., Ruangsomboon, O., Chuakul, W., Eampokalap, B., 1996. Serotypes of *Cryptococcus neoformans* isolated from patients prior to and during the AIDS era in Thailand. Mycopathologia 135, 75–78. https://doi.org/10.1007/BF00436454

Sun, S., Xu, J., 2009. Chromosomal Rearrangements between Serotype A and D Strains in *Cryptococcus neoformans*. PLoS One 4, e5524. https://doi.org/10.1371/journal.pone.0005524

Sun, S., Xu, J., 2007. Genetic analyses of a hybrid cross between serotypes A and D strains of the human pathogenic fungus *Cryptococcus neoformans*. Genetics 177, 1475–86. https://doi.org/10.1534/genetics.107.078923

Tanaka, R., Nishimura, K., Miyaji, M., 1999. Ploidy of serotype AD strains of

Cryptococcus neoformans. Nihon Ishinkin Gakkai Zasshi 40, 31–4.

Taylor, J.W., Fisher, M.C., 2003. Fungal multilocus sequence typing — it's not just for bacteria. Curr. Opin. Microbiol. 6, 351–356. https://doi.org/10.1016/S1369-5274(03)00088-2

Thompson, G.R., Wiederhold, N.P., Najvar, L.K., Bocanegra, R., Kirkpatrick, W.R., Graybill, J.R., Patterson, T.F., 2012. A murine model of *Cryptococcus gattii* meningoencephalitis. J. Antimicrob. Chemother. 67, 1432–1438. https://doi.org/10.1093/jac/dks060

Tomazin, R., Matos, T., Meis, J.F., Hagen, F., 2017. Molecular Characterization and Antifungal Susceptibility Testing of Sequentially Obtained Clinical *Cryptococcus deneoformans* and *Cryptococcus neoformans* Isolates from Ljubljana, Slovenia. Mycopathologia 1–10. https://doi.org/10.1007/s11046-017-0214-9

Van Wyk, M., Govender, N.P., Mitchell, T.G., Litvintseva, A.P., GERMS-SA, 2014. Multilocus sequence typing of serially collected isolates of *Cryptococcus* from HIV-infected patients in South Africa. J. Clin. Microbiol. 52, 1921–31. https://doi.org/10.1128/JCM.03177-13

Velagapudi, R., Hsueh, Y.-P., Geunes-Boyer, S., Wright, J.R., Heitman, J., 2009. Spores as infectious propagules of *Cryptococcus neoformans*. Infect. Immun. 77, 4345–55. https://doi.org/10.1128/IAI.00542-09

Venn-Watson, S., Daniels, R., Smith, C., 2012. Thirty year retrospective evaluation of pneumonia in a bottlenose dolphin Tursiops truncatus population. Dis. Aquat. Organ. 99, 237–242. https://doi.org/10.3354/dao02471

Viviani, M.A., Antinori, S., Cogliati, M. et al. 2002. European Confederation of Medical Mycology (ECMM) prospective survey of cryptococcosis: Report from Italy. Med. Mycol. 40.

Viviani, M.A., Cogliati, M., Esposto, M.C. et al. 2006. Molecular analysis of 311 *Cryptococcus neoformans* isolates from a 30-month ECMM survey of cryptococcosis in Europe. FEMS Yeast Res. 6, 614–619.

https://doi.org/10.1111/j.1567-1364.2006.00081.x

Viviani, M.A., Nikolova, R., Esposto, M.C., Prinz, G., Cogliati, M., 2003. First European case of serotype A *MAT**a** Cryptococcus neoformans* infection. Emerg. Infect. Dis. 9, 1179–80. https://doi.org/10.3201/eid0909.020770

Viviani, M.A., Wen, H., Roverselli, A., Caldarelli-Stefano, R., Cogliati, M., Ferrante, P., Tortorano, A.M., 1997. Identification by polymerase chain reaction fingerprinting of *Cryptococcus neoformans* serotype AD. J. Med. Vet. Mycol. 35, 355–360. https://doi.org/10.1080/02681219780001411

Voelz, K., Lammas, D.A., May, R.C., 2009. Cytokine signaling regulates the outcome of intracellular macrophage parasitism by *Cryptococcus neoformans*. Infect. Immun. 77, 3450–7. https://doi.org/10.1128/IAI.00297-09

Voelz, K., Ma, H., Phadke, S. et al. 2013. Transmission of Hypervirulence Traits via Sexual Reproduction within and between Lineages of the Human Fungal Pathogen *Cryptococcus gattii*. PLoS Genet. 9, e1003771. https://doi.org/10.1371/journal.pgen.1003771

Vogan, A.A., Khankhet, J., Samarasinghe, H., Xu, J., 2016. Identification of QTLs Associated with Virulence Related Traits and Drug Resistance in *Cryptococcus neoformans*. G3 Genes|Genomes|Genetics g3.116.029595. https://doi.org/10.1534/G3.116.029595

Vogan, A.A., Khankhet, J., Xu, J., 2013. Evidence for Mitotic Recombination within the Basidia of a Hybrid Cross of *Cryptococcus neoformans*. PLoS One 8. https://doi.org/10.1371/journal.pone.0062790

Vu, K., Tham, R., Uhrig, J.P., Thompson, G.R., Na Pombejra, S., Jamklang, M., Bautos, J.M., Gelli, A., 2014. Invasion of the central nervous system by *Cryptococcus neoformans* requires a secreted fungal metalloprotease. MBio 5, e01101-14. https://doi.org/10.1128/mBio.01101-14

Wendte, J.M., Miller, M.A., Lambourn, D.M., Magargal, S.L., Jessup, D.A., Grigg, M.E., 2010. Self-Mating in the Definitive Host Potentiates Clonal Outbreaks of the Apicomplexan Parasites Sarcocystis neurona and Toxoplasma gondii. PLoS

Genet. 6, e1001261. https://doi.org/10.1371/journal.pgen.1001261

Xu, J., Luo, G., Vilgalys, R.J., Brandt, M.E., Mitchell, T.G., 2002. Multiple origins of hybrid strains of *Cryptococcus neoformans* with serotype AD. Microbiology. https://doi.org/10.1099/00221287-148-1-203

Xu, J., Mitchell, T.G. 2003. Comparative gene genealogical analyses of strains of serotype AD identify recombination in populations of serotypes A and D in the human pathogenic yeast *Cryptococcus neoformans.* Microbiology. 149:2147-2154. doi: 10.1099/mic.0.26180-0

Xu, J., Mitchell, T.G., Litvintseva, A.P., 2011. Population Structure and Ecology of *Cryptococcus neoformans* and *Cryptococcus gattii*, in: *Cryptococcus*. American Society of Microbiology, pp. 97–111. https://doi.org/10.1128/9781555816858.ch8

Xu, J., Vilgalys, R., Mitchell, T.G., 2000. Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*. Mol. Ecol. 9, 1471–1481. https://doi.org/10.1046/j.1365-294x.2000.01021.x

Yan, Z., Li, X., Xu, J., 2002. Geographic distribution of mating type alleles of *Cryptococcus neoformans* in four areas of the United States. J. Clin. Microbiol. 40, 965–72. https://doi.org/10.1128/JCM.40.3.965-972.2002

Zaragoza, O., Chrisman, C.J., Castelli, M.V. et al. 2008. Capsule enlargement in *Cryptococcus neoformans* confers resistance to oxidative stress suggesting a mechanism for intracellular survival. Cell. Microbiol. 10, 2043–2057. https://doi.org/10.1111/j.1462-5822.2008.01186.x

Chapter 6

Hybridization Facilitates Adaptive Evolution in Two Major Fungal Pathogens

6.1 Preface

Hybridization is recognized as an important phenomenon that drives adaptation and evolution within the fungal kingdom. This is especially true for fungal pathogens of humans and animals where hybridization between closely related lineages can give rise to offspring with enhanced virulence. In this collaborative review article, we provided a comprehensive overview of the cryptococcal mating cycle, global epidemiology, and the genetic architecture of cryptococcal hybrids. Our collaborators synthesized the current state of knowledge on hybrid lineages of chytridiomycete frog pathogen *Batrachochytrium dendrobatidis.* We proposed that experimentation on *Cryptococcus* hybrids can be used as a model system for shedding light on less well understood fungal hybrids such as *B. dendrobatidis.*

As the first co-author, I was one of the two primary contributors of this work. With my co-author Man You, I reviewed the literature and drafted the sections on cryptococcal hybrids. Other co-authors of this article are Thomas Jenkinson, Jianping Xu and Timothy James. This manuscript is now published in Genes volume 11, issue 1.

6.2 Abstract

Hybridization is increasingly recognized as an important force impacting adaptation and evolution in many lineages of fungi. During hybridization, divergent genomes and alleles are brought together into the same cell, potentiating adaptation by increasing genomic plasticity. Here, we review hybridization in fungi by focusing on two fungal pathogens of animals. Hybridization is common between the basidiomycete yeast species *Cryptococcus neoformans* x *Cryptococcus* de*neoformans*, and hybrid genotypes are frequently found in both environmental and clinical settings. The two species show 10-15% nucleotide divergence at the genome level, and their hybrids are highly heterozygous. Though largely sterile and unable to mate, these hybrids can propagate asexually and generate diverse genotypes by nondisjunction, aberrant meiosis, mitotic recombination and gene conversion. Under stress conditions, the rate of such genetic changes can increase, leading to rapid adaptation. Conversely, in hybrids formed between lineages of the chytridiomycete frog pathogen *Batrachochytrium dendrobatidis* (Bd), the hybrids are considerably less diverged (0.2% divergent). The Bd hybrids are formed from crosses between lineages that rarely undergo sex. A common theme in both species is that hybrids show genome plasticity via aneuploidy and loss of heterozygosity and leverage it as a rapid way to generate genotypic/phenotypic diversity. Some hybrids show greater fitness and survival in both virulence and virulence-associated phenotypes than parental lineages under certain conditions. These studies showcase how experimentation in model species such as *Cryptococcus* can be a powerful tool in elucidating the genotypic and phenotypic consequences of hybridization.

6.3 Introduction

Hybridization refers to the interbreeding of individuals from genetically distinct populations or species. Through hybridization, genes and alleles that have diverged significantly from each other are brought together into the same cells and individuals, potentially creating the scenario of novel interactions among genes and genomes. There are two contrasting views on the roles of hybridization in organic evolution. On the one hand, hybridization is considered an evolutionary noise with limited long-term effect. Indeed, under a model in which species diverge by adaptation to different

niches, the majority of hybrids are expected to show lower fitness than either parental genotype in parental niches [1,2]. On the other hand, certain hybrid progeny may display transgressive segregation/ extreme phenotypes or fitness, both positive and negative, that exceed parental values, especially in novel ecological niches. Because of this, the contrasting view believes that hybridization plays critical roles in generating evolutionary novelty and can impact long-term evolution [3,4].

Hybrids may be either homoploid or polyploid. Homoploid hybrids have the same ploidy as the parents, but they may face obstacles to further sexual reproduction due to potential chromosomal incompatibilities. If homoploid hybrids can self-fertilize or backcross, the resulting offspring often reveal a high genetic and phenotypic variance, including the generation of transgressive phenotypes [5]. Polyploid hybrids, specifically allopolyploids, are generated typically by chromosome duplication either in gametogenesis or after zygote formation via mating of different species [6]. The presence of even numbers of homologous chromosomes from the same species alleviates the pairing problems in meiosis due to low sequence similarity or structural rearrangements that lead to hybrid sterility. Outcomes of hybridization are diverse. For example, in plants, hybridization may lead to the formation of novel hybrid species that are genetically isolated and phenotypically distinct from progenitors [3,5]. In other contexts, the extent of hybridization may be limited to a geographic region (or hybrid zone) maintained by either environment-genotype interactions or through a balance between migration and selection. In any case, hybridization makes possible the introgression of alleles from one parental species into another, which is becoming easier to identify through genome sequencing [7,8].

In all major groups of eukaryotes such as plants, animals, and fungi, natural hybridization has been reported, with some groups showing over 20% of extant species as hybrids. Previously, however, hybridization in fungi was either ignored or discounted, and only since the molecular revolution has hybridization in fungi been increasingly accepted as playing important roles [9,10]. Fungi are unique from the better-studied plants and animals in that they often produce copious (literally millions or billions) amounts of recombinants from a single mating event, which could lead to the generation of immense diversity. They are also unique because their dispersal is

not considered particularly limited, such that traditional hybrid zones of a limited geographic extent seem unlikely. On the other hand, many fungi display a mixed mode of reproduction, with extensive generations of asexual reproduction interspersed with rare sexual reproduction [11]. Such versatility in reproduction could allow hybrids to at least propagate asexually for extended periods of time without suffering from a potential segregation load [12]. In fungi, the F1 hybrids often display aneuploidy, diploidy, or higher ploidy, and because of the extra chromosome copies, they can continue to diversify and adapt through mechanisms such as mitotic recombination and gain/loss of individual chromosomes.

In this review, we explore these issues with a fungal view of hybridization described using two major fungal pathogens, the human-pathogenic *Cryptococcus* species and the amphibian chytrid fungus *Batrachochytrium dendrobatidis*. We argue that hybridization in these two groups of fungi reveals much about how speciation and hybridization in fungi is unique, and how hybridization will become increasingly important to both the environment and human health in this era of rapid fungal evolution. The choice of these two species, obviously, is based off of the authors' experiences. However, the pair complement each other nicely. On the one hand, *Cryptococcus* (Basidiomycota) is a model system and easily manipulated in the lab, while *B. dendrobatidis* (Chytridiomycota), a species of great ecological significance, is far from a model system and lacks genetic tools. Insights into hybridization in these species require different approaches, but much has been revealed and facilitated by genomics.

6.4 Hybridization in *Cryptococcus* species complex

Cryptococcus species complex

The human pathogenic *Cryptococcus* species complex comprises a group of closely related, basidiomycetous yeast species, responsible for over 220 000 annual infections worldwide with a mortality rate of ~80% [13]. Cryptococcal infections, known to occur in multiple forms including respiratory infections, skin lesions, and meningoencephalitis, are collectively called cryptococcosis and are a leading cause of death among HIV/AIDS patients worldwide. Commonly found in association with soil, bird droppings and tree barks, *Cryptococcus* species has a global distribution with

strains having been discovered from all continents except Antarctica [14]. Currently, seven evolutionarily divergent lineages are recognized as pathogenic *Cryptococcus* species, which can be distinguished based on genetic and molecular characterization, with each species having been assigned specific molecular types (**Figure 1**) [15,16]. Historically, *Cryptococcus* strains were categorized into serotypes based on the structure of polysaccharides at the cell surface. An alternative classification system aims to maintain two major species that are subdivided into varieties and/or molecular types [17,18]. In this review, we will use the seven-species classification system with (i) *C. neoformans* (serotype A, molecular types VNI, VNII and VNB; VN = variety *neoformans*), (ii) *C. deneoformans* (serotype D, molecular type VNIV), (iii) *C. gattii* (molecular type VGI; VG = variety *gattii*), (iv) *C. bacillisporus* (molecular type VGIII), (v) *C. deuterogattii* (molecular type VGII), (vi) *C. tetragattii* (molecular type VGIV) and (vii) *C. decagattii* (molecular type VGIV/VGIIIc) (the last five species share serotypes B and C). To maintain consistency with previous literature, we will categorize cryptococcal hybrids based on the serotypes of the parental species. For example, hybrids of *C. neoformans* and *C. deneoformans* will be referred to as AD hybrids.

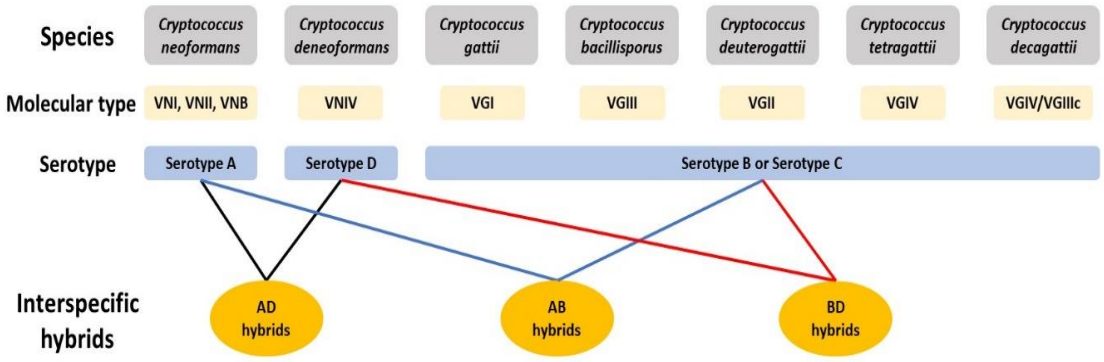


Figure 1: The currently recognized species in the pathogenic *Cryptococcus* species complex. The seven species of this complex can be distinguished based on genetic and molecular differences, and each is assigned a distinct molecular type. Historically, cryptococcal strains were broadly categorized into serotypes based on the antigens found at the cell surface. Hybrids arising from mating between species are named based on the serotypes of the parental strains.

The seven species are believed to be descended from a common ancestor that existed on the supercontinent Pangea. It is hypothesized that the ancestral population was split into two, distinct groups following the breakup of Pangea and subsequent continental drift further spatially isolated the subgroups, leading to the emergence of two major, independently evolving populations in South America and Africa respectively. This theory is supported by the fact that the estimated split of VN molecular types from VG molecular types occurred ~100 million years ago (mya) which coincides with the breakup of Pangea [19]. Furthermore, genomic analyses have revealed the origin of VN molecular types to be Africa while VG molecular types originated in South America. The ancestral populations of VN and VG molecular types continued to diverge and evolve, likely due to local niche differences, eventually splitting into the seven species observed today. *C. neoformans* and *C.* d*eneoformans* are believed to have diverged ~24.5 mya [19–22]. More recently, within C. *neoformans*, VNI and VNII split from each other ~4.7 mya. Among VG lineages, VGII diverged from other VG lineages ~12.5 mya, followed by VGIV splitting ~11.7 mya. VGI and VGIII were the last to diverge from each other approximately 8.5 mya [21].

Over the last century, historical spatial barriers between the seven cryptococcal species have been challenged with the significant increase in international commercial and human travel. The different lineages are now thrust back into contact due to anthropogenic transfer of cryptococcal cells/strains across countries and continents. Despite significant genomic nucleotide divergence, these species are still sufficiently compatible to initiate mating with each other, making interspecific hybridization a significant force that shapes their ongoing evolution. In fact, cryptococcal hybrids with superior fitness to parental strains have been recovered from both natural and laboratory settings while the proportion of infections caused by C. *neoformans* x C. *deneoformans* (AD) hybrids is on the rise, especially in Europe where these hybrids are responsible for nearly 40% of all cryptococcal infections. The implications of interspecific hybridization on the adaptive evolution of pathogenic *Cryptococcus* species complex are discussed below.

Sexual cycle of *Cryptococcus*

Cryptococcus species are haploid basidiomycete yeasts that can reproduce asexually via budding or sexually via mating (Figure 2). The sexual cycle of *Cryptococcus* was first observed by Kwon-Chung four decades ago [23,24]. Under nutrient-limiting conditions (e.g., low nitrogen) and dehydration, cells of opposite mating types (*MAT***a** and *MAT***α**) could be triggered to fuse and form a filamentous dikaryon with the two parental nuclei inside. A germ tube originating from the *MAT***α** end of the zygote extends out and subsequently develops into dikaryotic hyphae. Septa form to separate the cells. During hyphal growth, clamp connections form across septa and fuse with the subapical neighbor cell. Haploid blastospores can form along the hyphae: blastospores are vegetative yeast-like cells that bud from the hyphae. Some hyphal cells can also enlarge and form chlamydospores which are thick-walled vegetative cells with condensed cytoplasm. Chlamydospores can facilitate long-term survival of cells in harsh environments [25]. During hyphal extension, one of the two daughter nuclei in the apical cells are transferred to subapical cells via clamp connections. Therefore, each hyphal cell is able to maintain two parental nuclei. At the onset of meiosis, the tip of an aerial hypha enlarges to form a basidium within which the two parental nuclei fuse and meiosis occurs to produce four recombinant, haploid nuclei. The daughter nuclei undergo repeated mitotic divisions with individual nuclei packaged into individual basidiospores that bud off from the basidium, consequently forming four chains of basidiospores [26–28].

Sexual mating in *Cryptococcus* is normally restricted to cells of opposite mating types. However, cryptococcal cells of the same mating type could be induced to mate with each other under nutrient-limiting conditions, referred to as same-mating-type reproduction [29–31]. Mating between genetically different strains of the same mating type can produce a high frequency of recombinants, similar to opposite-mating-type mating [27]. Even though both *MAT***a** and *MAT***α** cells can undergo same-mating-type reproduction, the presence of cells with opposite mating types significantly enhances this process [32]. *MAT***α** is the predominant mating type in natural populations, with >99% of *C. neoformans* strains and ~95% of other cryptococcal strains belonging to this mating type [14,33,34]. During same-mating-type mating, haploid *MAT***α** cells

become **α**/**α** diploids, either by endoduplication or by nuclear fusion following cell-cell fusion between two **α** cells [27,35]. The newly formed diploid nucleus then undergoes one round of meiosis to produce four, recombinant haploid nuclei, which are then packaged into basidiospores [36]. Evidence for same-mating-type mating in *Cryptococcus* has also been found in nature: particularly, same-mating-type reproduction could be beneficial due to natural populations in both clinical and environmental settings being predominantly of the same mating type, *MAT***α** [37,38].

Both mating processes confer two major benefits to *Cryptococcus*: (i) it results in the production of basidiospores which act as the unit of long distance cryptococcal dispersal, and (ii) it is known to be mutagenic, generating both genetic and phenotypic variation as described in the sections below [39]. Strains of different cryptococcal species have been successfully mated with each other in the laboratory, albeit generally with low spore germination rates [40]. Unlike haploid parental strains, the resulting interspecific hybrids are often diploid or aneuploid, indicating that genomic divergence between parental lineages could hinder chromosomal disjunction and proper meiosis during gamete formation.

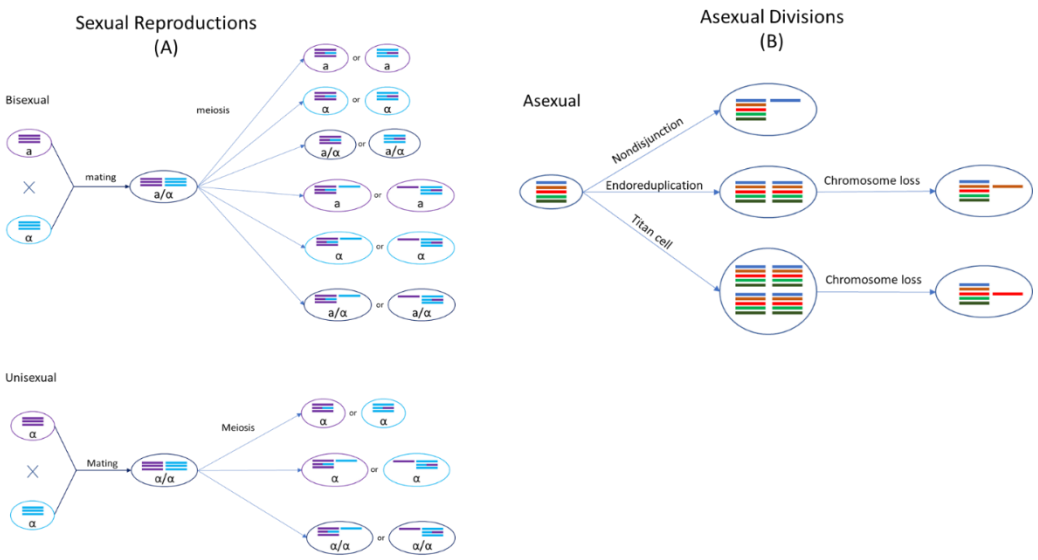


Figure 2: Genomic changes occur in sexual and asexual processes in *Cryptococcus*. (A) In sexual mating and subsequent meiosis, resulting in the formation of recombinant cells, both euploidy and aneuploidy can be observed; (B) Asexual replication can undergo nondisjunction, endoreduplication or the formation of titan cells to cause genomic changes.

Hybrids in the *Cryptococcus* species complex

Cryptococcal hybrids, discussed below, are categorized according to the serotypes of the parental lineages.

AB and BD hybrids: To the best of our knowledge, hybrids arising from mating of *C. neoformans* (serotype A) or *C. deneoformans* (serotype D) with one of *C. gattii, C. bacillisporus, C. deuterogattii, C. tetragattii or C. decagattii* (serotypes B or C) have not been recovered from the environment. However, there have been several reports of infections caused by AB (serotype A x serotype B) and BD (serotype B x serotype D) hybrid strains. The first such hybrid was reported in 2006 when three strains isolated from the cerebrospinal fluid of two patients in the Netherlands were determined to be diploid strains of serotype BD [44]. Based on these findings, it was estimated that BD hybrids make up about 1% of the clinical isolates found in the Netherlands between 1977 and 2005. A cryptococcal hybrid of serotype A x serotype B was first reported in 2008 when a strain isolated from an AIDS patient in Canada was determined to be an AB hybrid [45]. In 2012, four AB hybrid strains were identified among clinical samples obtained from the cerebrospinal fluid of patients in Brazil (n=2), Columbia (n=1) and India (n=1) [46]. AB and BD hybrids have since been reported in Germany [47], Denmark [48], and the United States [49], all from clinical sources. Even though mating between the parental lineages can be induced in the laboratory, the mating cells often fail to complete the sexual cycle, indicating a strong post-zygotic reproductive barrier [24]. These observations suggest that while these species are still genetically and phenotypically compatible enough to initiate mating, they produce relatively few viable hybrid progeny, leading to their rare occurrence in environmental and clinical samples. Furthermore, as of now, cryptococcal hybrids of serotypes AC and CD have never been reported in the literature.

AD hybrids: Hybridization between A and D serotypes is a significantly more common occurrence than hybridization of A or D serotypes with B or C serotypes. Indeed, *C. neoformans* x *C. deneoformans* hybrids, commonly referred to as AD hybrids, are the most common of all cryptococcal hybrids. AD hybrids are assigned their own molecular type of VNIII. Since their initial discovery in 1977, their prevalence has been steadily increasing, with AD hybrids currently causing up to 40%

of all cryptococcal infections in Europe [50–53]. They have been recovered from clinical and environmental sources in many countries across continents [54]. AD hybrids are often diploid or aneuploid and experience frequent loss of heterozygosity during vegetative growth [41,55,56].

BC hybrids: Under mating-inducing conditions, strains of different VG molecular types (serotypes B and C) are capable of mating with each other in the laboratory [40,57–59]. However, to the best of our knowledge, diploid/aneuploid VG hybrid strains (BC hybrids) have not been recovered from environmental or clinical sources. One likely explanation might be that VG lineages are still sufficiently compatible with each other to produce haploid, recombinant progeny: nucleotide divergence between VG lineages is less extensive in comparison to that between VNI and VNIV (see section 1.1). In fact, when strains of serotypes B and C mated with each other, they produced proper dikaryotic hyphae with clamp connections that were morphologically similar to those produced when VG strains of the same serotype mated with each other [24,57]. However, the spore viability was found to be very low (< 1%) in lab-derived reciprocal crosses of VGII (*MAT***α**) x VGIII (*MAT***a**) and VGII (*MAT***a**) x VGIII (*MAT***α**): among the spores that successfully germinated, 18/18 and 9/16 were diploid/aneuploid in the two crosses respectively [60]. Furthermore, almost all collected F1 hybrid progeny were determined to be diploid/aneuploid when VG lineages were mated with each other, or VGIII was mated with VN lineages in the laboratory (You et al., unpublished data). Together, these observations suggest that VG hybrids can be produced, and the apparent absence of diploid/aneuploid VG hybrids in nature is likely due to these hybrids' failure to successfully compete with parental lineages in their natural habitats.

6.5 Outcomes of hybridization between cryptococcal lineages

Hybrid inviability

Spore germination rates in hybrid cryptococcal crosses are typically low, indicating significant post-zygotic reproductive isolation between these species. Studies have reported ~5-20% of hybrid spores to be viable in lab-derived hybrid crosses between *C. neoformans* and *C. deneoformans* [41,61,62]. However, the true germination rate might be slightly higher since we have observed some AD hybrid spores to display an

abortive phenotype where following germination, growth is aborted after several mitotic divisions, indicating genetic incompatibilities within the nucleus [61]. Two recent studies detected significant variability in spore germination rates in crosses between strains of *C. gattii, C. bacillisporus, C. deuterogattii, C. tetragattii and C. decagattii* with a range of ~1-98% [29,40,60]: however, since the ploidy of germinated spores was not determined in the study by You et al., the proportion of viable progeny that were diploid/aneuploid is not known. You et al. also found spore viability to vary between ~1-43% in a series of crosses where a *Cryptococcus bacillisporus* strain was crossed with different C. *neoformans* and *C. deneoformans* strains [29]. The significant variation in spore viability observed across multiple studies highlights the complex interplay of determinants, including parental genetic backgrounds as well as environmental and genotype-environment interaction effects, on the germination of hybrid cryptococcal spores [29,58,59].

Hybrid sterility

While most cryptococcal hybrids are heterozygous at multiple loci across the genome and contain alleles for both mating types (i.e. heterozygous at the MAT locus), few are self-fertile or can mate with other strains [41,63]. Basidiospores produced by three self-fertile AD hybrids containing both mating types germinated at a very low rate of ~5% in laboratory conditions: these three AD hybrids did not produce any sexual spores when co-incubated with haploid *MAT***a** and *MAT***α** strains of *C. neoformans* and *C. deneoformans* [41]. However, sterility does not pose a barrier to cryptococcal hybrid success as they can propagate asexually via mitosis.

Phenotypic diversity and hybrid vigor

Studies investigating phenotypes and virulence of cryptococcal hybrids have found variable results with some reporting hybrid vigor (heterosis) while others have found hybrids to be inferior to parental strains. In general, AD hybrids are less virulent than either C. *neoformans* or C. d*eneoformans* haploid strains, though this result is not always observed [64]. However, increasing evidence of hybrid vigor in both natural and laboratory-constructed AD hybrids has been found. For example, AD hybrids might be better able to adapt to new environmental niches than C. *neoformans* and *C.*

deneoformans isolates. Natural AD hybrid strains are more resistant to UV irradiation than native C. *neoformans* strains from Botswana [65]. Laboratory-constructed AD hybrids also showed hybrid vigor with higher resistance to both UV irradiation and high temperatures than C. *neoformans* and *C. deneoformans* parents [65,66]. Another study found the majority of 31 investigated global AD isolates to be resistant to the antifungal drug FK506 [56,67]. A small proportion of hybrid strains in an AD hybrid population derived from a cross between CDC15 (C. *neoformans*, MAT**α**) and JEC20 (C. *deneoformans*, *MAT***a**) was found to surpass both parents in the expression of essential virulence factors including melanin production, capsule production, growth at 37℃, resistance to the antifungal fluconazole, and cell size [68]. The remaining hybrid offspring displayed intermediate phenotypes or inferior phenotypes to both parents. At present, similar phenotypic data on other cryptococcal hybrids is not available due to their rarity compared to AD hybrids. However, the presence of these hybrids in clinical settings suggests that at least some are capable of causing fatal infections in humans.

The phenotypic diversity found among cryptococcal hybrids indicates the presence of extensive genetic diversity in hybrid populations. Due to significant genomic differences between the parental species, frequent chromosome nondisjunction is observed during meiosis, with hybrids often inheriting novel and/or unique combinations of chromosomes. For example, evidence of homozygosity (or hemizygosity) interspersed with heterozygosity is observed across AD hybrid genomes [61,68,69]. Homozygosity could be derived through either chromosome loss or mitotic gene conversion leading to loss of heterozygosity [56,67,70]. Generation of novel allelic and chromosomal combinations can offer cryptococcal hybrids a significant advantage in adapting to a diversity of environmental niches and competing with parental lineages for resources.

6.6 Genetics of cryptococcal hybrids

Aneuploidy in cryptococcal hybrids

AD hybrids are either diploid or aneuploid as determined by fluorescence-activated cell sorting (FACS) analysis [41,52]. Previous studies revealed that aneuploidy in AD hybrids is most likely caused by non-disjunction of homologous chromosomes during

meiosis due to nucleotide sequence divergence (10%-15%), as well as genetic incompatibilities, between C. *neoformans* and C. de*neoformans* genomes [41,71,72]. Sun and Xu found that at least one out of 114 screened co-dominant loci was heterozygous in the majority of lab-derived AD hybrid offspring strains with an average heterozygosity of ~75% per strain [55]. Recombination between markers located on the same chromosome was observed, confirming the involvement of a meiotic process in the generation of these progeny, although the rate of crossovers was significantly lower during hybridization than that observed in intraspecific crosses of C. *neoformans* and C. *deneoformans*. Another analysis of a hybrid cross between H99 (C. *neoformans, MAT***α**) and JEC20 (C. *deneoformans, MAT***a**) suggested that the resulting hybrid progeny were likely generated via random nuclear fusion of two of the four recombinant nuclei generated from meiosis, which could result in heterozygous hybrids with doubled ploidy levels [73].

Clinical BD hybrid strains have also been found to be diploid or aneuploid by Bovers et al. [45]. These hybrids have a unique Amplified Fragment Length Polymorphism (AFLP) genotype (AFLP genotype group 8), revealing that they likely originated from hybridization between a *MAT***α**, serotype B strain (AFLP genotype 4) and a *MAT***a,** serotype D strain (AFLP genotype 2). Interestingly, these hybrids were heterozygous at two of the genotyped loci, namely RNA polymerase II (RPB2) and laccase (LAC): however, they were homozygous for the serotype B parent's genotype at the Internal Transcribed Spacer (ITS) regions of the nuclear ribosomal RNA gene cluster. In addition, most BD hybrids were homozygous for the serotype B allele at the Intergenic Spacer (IGS) sequence of the nuclear ribosomal RNA gene cluster, while the remaining hybrids were homozygous for the serotype D allele.

While diploid/aneuploid hybrid strains of *C. gattii, C. bacillisporus, C. deuterogattii, C. tetragattii and C. decagattii* have not been recovered from nature, lab-derived hybrids of such crosses often display diploidy or aneuploidy. In two laboratory crosses between *C. bacillisporus* and *C. deuterogattii*, 18/18 and 9/16 of the spores that successfully germinated were determined to be diploid/aneuploid respectively [60]. Furthermore, almost all collected F1 hybrid progeny were determined to be diploid/aneuploid when these five species were mated with each other, or when *C.*

bacillisporus was mated with C. *neoformans* and *C. deneoformans* in the laboratory (You et al., unpublished data).

Loss of heterozygosity

The definition of loss of heterozygosity (LOH) is the loss of one parental allele in a certain genomic region in a heterozygous individual. LOH can be caused by multiple mechanisms, such as unbalanced chromosome rearrangements, gene conversion, mitotic recombination, and loss of a chromosome or a chromosomal segment. Double-strand break repair can give rise to short-range LOH events by gene conversion without crossover. However, long-range LOH events are mostly caused by single crossovers or break-induced replication [74]. In addition, whole-chromosome LOH can arise by chromosome loss through nondisjunction, followed by duplication of the remaining homolog [75]. The duplication of a chromosome is a common occurrence in whole-chromosome LOH. With complete duplication of the remaining genetic material, the appearance of a normal karyotype is maintained, even though there may have been a wholesale loss of genetic diversity. Generally, LOH is not reversible, however, cells can regain the lost heterozygous alleles via outcrossing or mutation.

The emergence of LOH is considered a major mechanism of generating genetic diversity in populations of diploid heterozygous organisms. Unlike the parental haploid lineages, cryptococcal AD hybrids are often highly heterozygous, and may be prone to LOH, both during hybridization events and during asexual growth following germination [39,41,52,56,66]. During sexual mating, the two parental nuclei have been observed to fuse at earlier stages of sexual development (e.g. in the zygote or hyphae), providing opportunities for mitotic recombination to facilitate LOH at certain chromosomal regions before meiosis. A recent analysis of 297 lab-derived AD hybrid progeny strains generated from a single cross revealed the hybrids to experience extensive loss of chromosomes [76]. Both partial and complete chromosome loss and duplication have been observed in some AD hybrids. Partial chromosome loss may result in genome rearrangement or the formation of novel chromosomes through truncation or translocation [56]. Li et al. found that the progeny strain P5 (progeny of a self-fertile AD strain CDC228) partially lost some chromosomes (chromosomes 8 and 10 from serotype A parent) and completely lost some others (chromosomes 5 and

13 from C. *neoformans*, and chromosomes 3 and 12 from *C. deneoformans*), giving rise to a highly unique genome organization [56].

Interestingly, natural AD hybrids show preferential retention of specific alleles and chromosomes from one of the two parents, suggesting that those alleles may offer survival and growth benefits under specific conditions [70]. Allele distributions in the genomes of AD hybrids often show significant departures from Mendelian ratios with alleles of one parent preferred over that of the other at certain loci [76]. In fact, Samarasinghe et al. [76] found genome-wide allele distribution in 297 AD hybrids to be significantly skewed in favor of the *C. deneoformans* parent from which the hybrids inherited mitochondria. It is hypothesized that given the uniparental mitochondrial inheritance seen in cryptococcal species, hybrids prefer to retain chromosomes of mitochondria-donor parent to minimize incongruence between their mitochondrial and nuclear genomes.

A very recent study conducted by Dong et al. estimated rate of LOH during mutation accumulation in a laboratory-constructed diploid AD hybrid (CDC15 x JEC20) during mitotic divisions. They used 33 genetic markers located on 14 chromosomes to determine genome-wide allele distributions in the AD hybrid [77]. The parental haploid strain CDC15 (serotype A, *MAT***α**) is more resistant to fluconazole (**M**inimum **I**nhibitory **C**oncentration = 64µg/ml) than the other parent JEC20 (serotype D, *MAT***a**, MIC = 4µg/ml). Their findings showed that only a few LOH events occurred over 800 generations of propagation on nutrient-rich medium, at a rate of 6.44 x 10^{-5} LOH events per mitotic division. However, fluconazole exposure resulted in a dramatic increase in LOH rate which increased more than 50-fold, at two markers on Chromosome 1. Interestingly, Chromosome 1 contains two genes, *ERG11* (the fluconazole target gene) and *AFR1* (a major transporter for triazoles), both of which play major roles in the development of antifungal resistance in *C. neoformans* [68,78]. Here, the AD hybrid lost the fluconazole-susceptible allele of both genes inherited from JEC20 while maintaining the alleles from CDC15. In these evolved strains, the copy number of Chromosome 1 inherited from CDC15 also increased. Results from this study suggested that hybridization can facilitate the rapid adaptation of *Cryptococcus* to stressful environmental conditions.

Dynamic ploidy changes in *Cryptococcus*

Ploidy change is often associated with sexual reproduction. Fungal cells generally mate with cells of identical ploidy levels, resulting in intermediate sexual structures with double the genomic content. Subsequent meiosis reduces the DNA content by half, reinstating the original ploidy of the parental strains. Some fungi, like *Saccharomyces cerevisiae*, favor propagation in the diploid state while other fungi, like *Schizosaccharomyces pombe*, prefer to propagate in the haploid state with a transient diploid state, as is the case observed in *Cryptococcus* [43,79,80].

Cryptococcal cells isolated from clinical and environmental settings are normally haploid with 14 chromosomes. FACS analyses have revealed an appreciable proportion of AD hybrids to be diploid [41,49,81]. However, it is possible that the diploid AD hybrids are not heterozygous at all loci across the genome: the remaining copy of a chromosome is often duplicated following LOH events, maintaining diploidy. Environmental stress has been observed to induce chromosome mis-segregation causing chromosomal instability. For example, exposure to a high-dose fluconazole treatment can result in the amplification of Chromosome 1 in both haploid C. *neoformans* and AD hybrids [77,78]. Comparison of haploid and diploid *C. neoformans* cells found that haploid cells were generally more virulent than diploid cells in a murine inhalation model of cryptococcosis [66]. However, in a rabbit infection model, diploids displayed similar virulence levels to haploid forms [80]. Higher ploidy was found to be associated with larger cell size in C. *neoformans*. For example, titan cells that can grow up to an impressive 100 µm in diameter contain 16, 32, 64 or more copies of the genome [82]. The large size facilitates the survival of titan cells during infection by hindering ingestion by host macrophages and by imparting resistance to oxidative and nitrosative stresses [75].

However, in certain cases, increased ploidy has been shown to have modest detrimental effects on virulence in a murine inhalation model, growth at high temperature, and melanisation [66]. In addition, melanin production was found to be correlated with monosomy at chromosome 13, while disomic variants produced less melanin and were less virulent in mice in *C. neoformans* cells isolated from AIDS patients [83]. The plasticity of their genomes provides cryptococcal hybrids with the

flexibility to alter their ploidy, via chromosome loss/gain or duplication, which in turn promotes adaptation to a wide range of environmental conditions.

Cryptococcus as a model system for fungal hybridization

Hybrids face significant challenges on survival and functionality due to two divergent genomes residing in the same cell. In a process referred to as genome stabilization, hybrids eliminate unfavorable combinations of the two parental genomes via a variety of mechanisms including recombination, gene conversion and chromosome loss [78]. The rate at which genome stabilization is achieved in a hybrid may be related to the extent of divergence between the two parental genomes since incompatibilities between more differentiated genomes will be resolved faster within the hybrids. For example, aneuploidy is commonly found in hybrids derived from two divergent parents, while frequent LOH events can be viewed as a mechanism of achieving genome stabilization.

Chromosomal nondisjunction during meiosis coupled with LOH during vegetative growth leads to the creation of cryptococcal hybrids with novel and unique allelic combinations not found in parental species. The novelty and plasticity of their genomes have put cryptococcal hybrids at a unique position to dynamically adapt to novel environmental niches and compete with parental lineages in current habitats. In fact, hybrid vigor displayed by some hybrids in laboratory settings and the increasing presence of AD hybrids in clinical samples reveal the ability of cryptococcal hybrids to successfully adapt to their environments. In summary, the genomic plasticity likely facilitates rapid adaptation of hybrids to new environmental niches (e.g. harsh environments) or genetic perturbations.

In a world where frequent international commerce and human travel is blurring geographical and spatial boundaries, hybridization between closely related taxa is becoming a common occurrence in all kingdoms of life. The seven species of the *Cryptococcus* species complex provide an excellent model system for studying hybridization in human fungal pathogens. Many tools have been developed and optimized for the study of these yeasts including numerous molecular, phenotypic and genetic assays, well-characterized laboratory strains as well as mouse models of

cryptococcal infections that make in vivo experiments feasible. Furthermore, whole genome sequences of hundreds of isolates from various geographical origins are freely available on online databases (ex: NCBI, FungiDB) while gene editing in these species can now be carried out with high efficiency using especially adapted CRISPR-Cas9 [84]. Finally, experimenting on haploid yeasts such as *Cryptococcus* is more convenient, and the findings can be more generalizable compared to well-established diploid model organisms such as *Saccharomyces cerevisiae*. Insights gained from cryptococcal hybrid research can be used to guide research efforts on hybridization in lesser known pathogenic fungi such as *B. dendrobatidis*, discussed in the remaining sections of this review.

6.7 Hybridization in an Aquatic Chytrid Fungus Associated with Amphibian Declines

The amphibian chytrid *Batrachochytrium dendrobatidis*

The chytrid fungus Batrachochytrium dendrobatidis (Bd) is a broad host-range pathogen that is known to infect close to 700 species of frogs, salamanders, and caecilians worldwide [85]. Bd is now recognized as the major contributor to near-simultaneous amphibian population declines in the 1980s and 1990s that are correlated with arrival of the pathogen [86,87]. To date, Bd has been detected on every continent except Antarctica [88] and is most likely introduced from source populations in east Asia [89].

Bd is composed of at least four deeply divergent evolutionary lineages with varying geographical histories and virulence against hosts. The most widespread and well documented of these lineages is a globally-distributed, hypervirulent diploid genotype, BdGPL (Global Panzootic Lineage) [90]. The other, putatively less virulent lineages of Bd include: a Brazilian lineage endemic to the Atlantic Forest region of southern Brazil, BdBrazil (also known as BdBrazil/Asia2) [89,91]; an African lineage endemic to the Cape region of South Africa, BdCape [90]; and an endemic Asian lineage BdAsia1, believed to be closest to the source of origin for Bd diversity [89]. Recently, the existence of an additional endemic lineage BdAsia3 widespread throughout southeast Asia was reported from amphibian swabs [92].

Hybrids in *Batrachochytrium dendrobatidis*

Intraspecific hybrid strains resulting from outcrossing between parental genotypes of divergent lineages are rare, but known to occur. Within the divergent lineages of Bd, reproduction appears to be strictly asexual [93] with the exception of BdAsia1, which has a population signature of a highly recombining population [93]. Outcrossing and hybridization are only known to occur in secondary contact zones where divergent lineages have been brought into proximity by human activity. There are currently five hybrid Bd isolates reported in the literature from two hybrid zones. Three of these hybrid isolates were documented from the Atlantic Forest of Brazil within a narrowly restricted zone in the southern Brazilian state of Paraná [91,94,95]. This locality is one of the areas where the BdGPL and BdBrazil lineages overlap. The other two known Bd hybrids are described from the Eastern Cape Province of South Africa where BdGPL and BdCape overlap. Putative hybrid isolates were also identified from genotyping DNA from amphibian skin swabs, but these could also be explained by coinfection [92]. Unlike the narrow geographic range of the Brazilian hybrid zone, the two South African hybrid isolates were collected approximately 200 km apart from one another [89]. Additional regions where secondary contact could occur along with hybridization are Europe, western Africa, and Central America [89,92,95]. These regions are of interest with respect to hybridization because they harbor BdGPL and BdCape lineages which are known to hybridize. Evidence of hybridization derives from combination of otherwise lineage-specific alleles into the same genome, increased heterozygosity, and Bayesian admixture analyses [89,94,95]. It is unclear whether hybrids are favored by natural selection where they are created. In order to test this, it would be useful to return to regions where hybridization has occurred in order to attempt reisolation of the same hybrid genotypes and to estimate frequencies of hybrids related to parental species. Both measures can test whether hybrid genotypes are on the increase, which is predicted if they are favored by selection.

While clear genetic evidence of hybridization in Bd exists, mating and hybridization have not yet been observed in situ or in the laboratory. Likewise, specialized meiotic structures have never been reported for this species. The cellular process of hybridization in Bd is of special interest to understanding the dispersal ecology of this pathogen because sexual reproduction in related members of the Chytridiomycota

results in production of environmental-resistant resting spores which may facilitate environmental or long-range transmission. Alternative cellular mechanisms of outcrossing that do not involve meiosis have also been proposed. One alternative mechanism by which Bd may be outcrossing is through a parasexual cycle [95]. Parasexual reproduction is well documented in other groups of pathogenic fungi, such as Candida albicans [96,97] This mode of reproduction involves the fusion of diploid cells without meiosis. The resulting cell is a tetraploid intermediate which in most cases loses chromosomal copies back to a diploid state (**Figure 3)**. Such a reproductive mechanism may explain the varying levels of aneuploidy prevalent throughout individual Bd isolates as well as the lack of obvious meiotic structures or resting spores in this species. On the other hand, given the diversity of mechanisms possible to create aneuploidy as discussed above for *Cryptococcus*, it is plausible that the sexual cycle of Bd is typical for other fungi, cryptic as it may be.

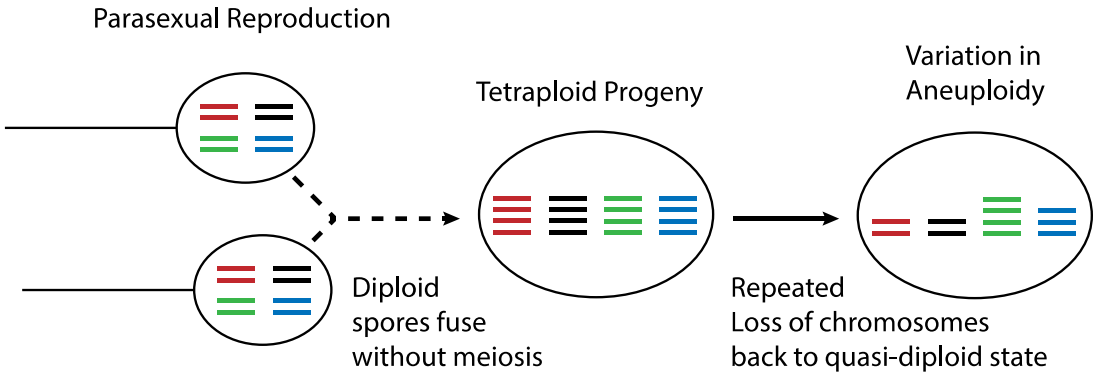


Figure 3: Parasexuality is a process of reproducing without a reductive cell division (meiosis). With parasexuality, a tetraploid offspring is produced which is a transient stage followed by random loss of chromosomes during vegetative growth.

Outcomes of hybridization in *Batrachochytrium dendrobatidis*

The phenotypic outcomes of hybridization in Bd remain largely unknown. In the only currently available study on hybrid phenotypes, Greenspan et al. [98] showed that hybrid virulence and pathogenicity was highly dependent on the infected host species. In a virulence challenge assay the authors infected two endemic Brazilian, direct-developing frog species with BdGPL, BdBrazil, or hybrid isolates produced by the

two lineages. In one host species, the high-altitude endemic pumpkin toadlet (Brachycephalus ephippium), hybrid isolates were more virulent – causing greater mortality in host animals – than either BdGPL or BdBrazil isolates. The endemic BdBrazil was the least virulent in this host species. In the other host species tested, the robber frog (Ischnocnema parva), hybrid isolates displayed an intermediate degree of virulence with BdGPL being the most, and BdBrazil being the least virulent.

Hybrid pathogenicity also varied according to host context. In the Greenspan et al. [98] study, the authors examined pathogenicity among isolates in three host species by assessing pathogen load upon host mortality. Again, the pathogenicity phenotypes of hybrids depended on the host species. In the host species, B. ephippium, where hybrid isolates were most virulent, the hybrids were also more pathogenic, with hybrid strains producing the highest spore loads on hosts at the time of mortality. In I. parva, the host species in which hybrid virulence was intermediate between BdGPL and BdBrazil, spore loads produced by hybrid isolates were comparable to those of BdGPL. The third species examined in the Greenspan et al. [98] study, the habitat-generalist, swamp treefrog (Dendropsophus minutus), is known to be more tolerant to Bd infection in laboratory challenge assays [99]. Because of this, infection experiments did not produce sufficient mortality in this species to analyze differences in virulence. However, the authors showed that Bd pathogenicity varied by genotype in this host species. Hybrid isolates produced spore loads intermediate between BdGPL (highest loads) and BdBrazil (lowest) when the host animals were assessed 60 days post-inoculation.

In summary, much work remains to be completed in understanding the cellular processes and phenotypic outcomes of hybridization in Bd. However, the Greenspan et al. study documents two very interesting results. First, hybrid genotypes do show some degree of heterosis (hybrid vigor) in both virulence and pathogenicity characteristics. Second, and very importantly, this increase in virulence and pathogenicity of the hybrid strains are context dependent on the host species being infected. These results highlight that predicting the phenotypic characteristics of hybrid Bd genotypes is very complex, and further experimental work should focus on assessing hybrid characteristics in a broad range of host species.

Aneuploidy in *Batrachochytrium dendrobatidis*

Although most fungi are considered to have zygotic meiosis where the diploid stage is highly limited, this is a gross generalization. In fact, there are no extensive studies of the genetics of Chytridiomycota that would allow ploidy changes during the life cycle to be fully elucidated. In Bd, an outcome of hybridization is the creation of genomes that are approximately 0.233% heterozygous [94]. This level of divergence is considerably lower than other hybrids like the cryptococcal hybrids and presumably should not provoke chromosome incompatibilities. Nonetheless, the presence of chromosomal rearrangements across Bd lineages has not been explored. Like *Cryptococcus*, Bd is also capable of generating genetic diversity via dynamic genome mutations including aneuploidy and LOH. While the Bd genome is generally considered to be diploid, chromosomal copy number can vary greatly. Whole genome sequencing studies of Bd find high variation in aneuploidy within individuals and among closely related isolates. Chromosomal copy number in Bd can vary between 1 (monosomic) to 5 (pentasomic). In 51 Bd genomes that have been sequenced and analyzed for copy number variation, two studies together show that approximately 58.7% of Bd chromosomes are disomic, 29.0% are trisomic, 11.3% are tetrasomic, and less than 1% monosomic or pentasomic [94,100].

The widespread nature of aneuploidy across major lineages of Bd suggests that it may be an important mechanism in the generation of genetic variation, especially given the rarity of sexual reproduction observed in this species. Links between variation in chromosomal copy number and phenotypic effect, however, have been difficult to establish in this species. This is largely due to the paucity of phenotyped Bd isolates with accompanying whole genome sequences. Comparatively, no obvious patterns have emerged to link highly aneuploid genomes to the hypervirulent BdGPL clade. The BdGPL lineage contains representative isolates displaying disomic, trisomic, and tetrasomic genomes throughout its phylogeny [94,100]. While the sample sizes of analyzed genomes outside the better-studied BdGPL clade are very small, some phylogenetic trends are beginning to emerge around Bd aneuploidy with respect to lineage. In the novel, enzootic Bd lineages, the BdCape isolates analyzed so far (n = 5) show a mix of trisomic (60%) and tetrasomic (40%) genomes [100]. Three

representative isolates assigned to BdCH were mostly trisomic. BdBrazil isolates (n = 2) were mostly disomic [94]. Finally, the single BdGPL/BdBrazil hybrid analyzed is also mostly disomic, with 13/17 of its major chromosomes disomic and 4/17 trisomic [94].

In addition to observations of widespread aneuploidy in Bd, chromosomal copy number has also been shown to be mutable over short timescales. Farrer et al. [100] examined replicate laboratory lines of an ancestrally trisomic Bd isolate serially passaged over 40 passages under differing growth conditions. One line was passaged in a standard media while the other was passaged in a selective media containing defensive antimicrobial peptides collected from the European water frog (Pelophylax esculentus) [101]. After approximately 40 weekly passages, the culture sequenced from standard media lost a copy of one chromosome (supercontig IV) and gained a copy of another (supercontig V), while the culture passaged in antimicrobial peptide media gained a chromosomal copy at supercontig V. Another study investigated genomic changes of an isolate, JEL427, before and after 30 transfers in the lab [102]. The isolate showed lower virulence and spore production after the passages [103]. The major change in the isolate over the 30 passages was a reduction in ploidy, going from an average of 3.6 copies per chromosome (in practice genomic scaffold as the genomic map is not available) to 3.1. The difference between the two laboratory evolution studies may relate to their divergent starting points, but both results show that aneuploidy changes can occur rapidly in Bd, perhaps serving as a mechanism for rapid genomic adaptation to changing selective pressures. While the functional underpinnings of these pattern are difficult to tease out, they may suggest that specific Bd chromosomes more readily gain or lose copies. This pattern is also reported in another major study investigating aneuploidy in Bd, where the authors find that supercontig V is one of the Bd chromosomes more likely to present higher than average copy numbers [94].

Characterization of the ploidy of additional hybrid isolates will be essential for understanding whether meiosis and sex result from fusion of haploid gametes or through a parasexual cycle. The general reduction of chromosome number over time from an ancestor of higher than disomic average chromosome numbers would tend to

support parasexuality, however, the absence of more than two alleles per locus suggests that more likely an endoduplication of all or part of the genome occurs sometime after hybridization. Either way, hybrid genotypes present a greater deal of allelic diversity which could facilitate adaptation by LOH. Understanding whether LOH occurs primarily at the chromosome level, i.e., aneuploidy, or at the gene level, e.g., gene conversion, in hybrid isolates is critical for understanding the nature and magnitude of the selective forces they experience.

LOH in *Batrachochytrium dendrobatidis*

Loss of heterozygosity is a well-known feature contributing to genome diversity in Bd. LOH is hypothesized to occur during asexual reproduction of Bd through mitotic recombination, chromosome loss, or gene conversion [93]. In addition to changes in chromosome copy number, LOH has great capacity to generate genotypic diversity without the input of new alleles [104]. This may be particularly important in clonally dominant pathogen lineages such as in Bd, such as the BdGPL lineage which typically shows a maximum of two alleles per locus, despite being frequently trisomic. The genotypic diversity generated by LOH should alter combinations of alleles within and across loci displaying overdominance, underdominance, or epistasis, which can then be subject to selection pressure with presumably advantageous tracts of LOH sweeping to fixation in a population.

The most prominent example of this in Bd may be a large, shared LOH region on supercontig II that is present in all members of BdGPL [89,94,100]. The conserved nature of this LOH feature throughout the globally invasive clade may reveal clues to the successful proliferation of this lineage. Gene functions enriched in shared LOH regions of BdGPL included processes related to reactive oxygen metabolism, L-serine metabolism, and superoxide dismutase/oxidoreductase [94]. These gene classes, in addition to various peptidases identified through comparative genomics with the closely related non-pathogenic chytrid Homolaphlyctis polyrhiza [105], and transcriptomic studies of Bd infections [106,107], are possibly involved in the genomic evolution underpinning the adaptive success of BdGPL in varied habitats worldwide.

In clonally reproducing diploid organisms such as Bd, the lack of outcrossing results in regions of LOH in the genome persisting in lineages through time. Because of this, shared homologous regions of LOH can be a powerful tool to inform populations of closely related clonal strains. For example, homologies in patterns of LOH have been used to distinguish evolutionary subclades within the global BdGPL [93,104]. Further geographic sampling and improved computational methods to detect homologies in LOH patterns hold the promise of resolving finer-scale intra-lineage population structure, and providing a more refined picture of the geographic history of this ecologically important pathogen.

6.8 Conclusions and Perspectives

Hybridization and genomic plasticity appear to be shared hallmarks contributing to rapid adaptation in fungal pathogens. The shared mechanisms of hybridization, aneuploidy, and LOH between the two major fungal pathogens *Cryptococcus* and *B. dendrobatidis* span the fungal tree of life (phylum Basidiomycota to Chytridiomycota), and appear to reflect a global pathway for rapid adaptation in pathogens across the fungal kingdom. Themes emerging from multiple studies of fungal hybrids are that they demonstrate increased ploidy, heterosis, and novel ecological niches. Prime examples of successful hybrids with these traits across the fungal tree of life are readily found. For example, hybridization in the plant endophytic species *Epichlöe* leads to asexual diploids/polyploids with major benefits to the host and hence fungus [108]. Likewise for the pathogen, *Verticillium longisporum*, hybridization is associated with expanded host range and diploidization [109,110]. Ancient hybridization is involved in the diversification of *Saccharomycete* yeasts [111], while more recent hybridization and polyploidization is involved in formation of yeasts (*Saccharomyces* spp.) involved in beer brewing [112,113]. Finally, two examples of human pathogenic species appear to be largely of hybrid origin: the halophilic black yeast *Hortaea warnecki* is the cause of superficial skin infection, tinea nigra [114] and the recently described, but rare, *Candida metapsilosis* [115].

The above-mentioned fungi generally are discovered as F1 hybrids with higher ploidy than their parental species or lineages. For example, in the case of *Cryptococcus*, like

most of the hybrids mentioned above, the hybrid isolates are generally diploid or aneuploid. On the other hand, for Bd, there is no evidence that F1 hybrids have higher ploidy than the parental genotypes and there is some evidence that some hybrids may be backcrossed or selfed (Jenkinson, unpublished). Other modes exist, such as homoploid hybrids in *Zymoseptoria pseudotritici* and *Microbotryum* spp. [116,117]. Following the formation of F1 hybrids, persistence and adaptation is facilitated by the mechanisms of genomic plasticity during asexual growth such ploidy cycling and LOH. The high number of successful asexual F1 hybrids with heterosis across species suggests that any reproductive isolation caused by genetic incompatibilities are likely to be recessive as predicted by theory [1]. Overall, however, it is plausible that the importance of hybridization in fungal adaptation may be overinterpreted as the likelihood of identifying or observing failed hybrids is low. The community would benefit from additional work synthesizing hybrids in the lab to understand and predict the outcome of hybridization on evolution, such as conducted for investigating reproductive isolation in *Neurospora* and *Microbotryum* [118,119]. More needs to be known about fungal adaptation to ecological gradients in order to understand whether hybrid zones are likely to exist at the boundaries in which allopatric species meet.

Understanding the evolutionary dynamics of better studied pathogens such as *Cryptococcus* can inform the biology of less characterized, or newly discovered fungal pathogens such as Bd. For example, in *Cryptococcus* most natural hybrids appear to be close to F1, and these data also agree with expectations of chromosomal pairing problems reflected in the 100 million year divergence between C. *neoformans* and *C. deneoformans* [19–22] and the studies which show meiotic segregation in F1 hybrids to be highly abnormal [55]. Though the time frame of divergence between the lineages of Bd is debated [89,94], at the extreme end an estimate of 100,000 years of divergence before hybridization is three orders of magnitude younger than cryptococcal hybrids. These data predict that F2 and further generation of hybrids of Bd are likely to exist, and it would be appropriate to account for this possibility when additional sampling in regions of admixture in conducted. As another example, mitochondrial genotype in *Cryptococcus* is not known to have an impact on hybrid fitness, yet this has been demonstrated in other species of fungal hybrids, using

crossing designs that create identical nuclear genomes in alternative mitochondrial association [80,120]. More work should be done to develop a crossing system in Bd that could allow mitochondrial-nuclear genetic interactions to be tested. Finally, in *Cryptococcus*, same-mating-type mating was first demonstrated to occur in nature, because the mating type locus had been well described and easily genotyped [37]. This is a large hurdle for a species like Bd, but the fact that Bd hybrids exist should provide impetus to determine the genetic basis of mating types which may reveal similar same-mating-type mating dynamics.

Regardless of underlying genetic differences between *Cryptococcus* and Bd, there are several commonalities regarding hybridization. Hybrids in both genera show increased genomic plasticity with the potential to generate transgressive phenotypes, unleashed following both meiotic and mitotic recombination. Hybrids in both genera reveal that it is a subset of environments, rather than all, in which they have higher fitness. Finally, the spread of both pathogens likely involves human-assisted migration that led to subsequent admixture. Given the potential adaptive benefits of hybridization in both *Cryptococcus* and Bd, improved measures for pathogen containment to prevent increased opportunities for hybridization should be put in place, both in clinical and environmental settings.

6.9 References

1. Barton, N.H. The role of hybridization in evolution. Mol. Ecol. **2001**, 10, 551–568.
2. Barton, N.H.; Hewitt, G.M. Analysis of Hybrid Zones. Annu. Rev. Ecol. Syst. **1985**, 16, 113–148.
3. Rieseberg, L.H. Hybrid Origins of Plant Species. Annual Review of Ecology and Systematics **1997**, 28, 359–389.
4. Arnold, M.L.; Hodges, S.A. Are natural hybrids fit or unfit relative to their parents? Trends Ecol. Evol. (Amst.) **1995**, 10, 67–71.
5. Mallet, J. Hybrid speciation. Nature **2007**, 446, 279–283.
6. Ramsey, J.; Schemske, D.W. Neopolyploidy in Flowering Plants. Annual Review of Ecology and Systematics **2002**, 33, 589–639.
7. Harrison, R.G.; Larson, E.L. Hybridization, introgression, and the nature of species boundaries. J. Hered. **2014**, 105 Suppl 1, 795–809.
8. Dunn, B.; Richter, C.; Kvitek, D.J.; Pugh, T.; Sherlock, G. Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. Genome Res **2012**, 22, 908–924.
9. Schardl, C.L.; Craven, K.D. Interspecific hybridization in plant-associated fungi and oomycetes: a review. Mol. Ecol. **2003**, 12, 2861–2873.
10. Stukenbrock, E.H. The Role of Hybridization in the Evolution and Emergence of New Fungal Plant Pathogens. Phytopathology **2016**, 106, 104–112.
11. Nieuwenhuis, B.P.S.; James, T.Y. The frequency of sex in fungi. Philos. Trans. R. Soc. Lond., B, Biol. Sci. **2016**, 371, 20150540.
12. Haag, C.R.; Roze, D. Genetic Load in Sexual and Asexual Diploids: Segregation, Dominance and Genetic Drift. Genetics **2007**, 176, 1663–1678.
13. Rajasingham, R.; Smith, R.M.; Park, B.J.; Jarvis, J.N.; Govender, N.P.; Chiller, T.M.; Denning, D.W.; Loyse, A.; Boulware, D.R. Global burden of disease of HIV-associated cryptococcal meningitis: an updated analysis. Lancet Infect Dis **2017**, 17, 873–881.
14. Cogliati, M. Global Molecular Epidemiology of *Cryptococcus neoformans* and *Cryptococcus gattii*: An Atlas of the Molecular Types. Scientifica **2013**, 2013, 1–23.
15. Hagen, F.; Khayhan, K.; Theelen, B.; Kolecka, A.; Polacheck, I.; Sionov, E.; Falk, R.; Parnmen, S.; Lumbsch, H.T.; Boekhout, T. Recognition of seven species in the *Cryptococcus gattii*/*Cryptococcus neoformans* species complex. Fungal Genet. Biol. **2015**, 78, 16–48.

16. Hagen, F.; Lumbsch, H.T.; Arsic Arsenijevic, V.; Badali, H.; Bertout, S.; Billmyre, R.B.; Bragulat, M.R.; Cabañes, F.J.; Carbia, M.; Chakrabarti, A.; et al. Importance of Resolving Fungal Nomenclature: the Case of Multiple Pathogenic Species in the *Cryptococcus* Genus. mSphere **2017**, 2, e00238-17.
17. Kwon-Chung, K.J.; Bennett, J.E.; Wickes, B.L.; Meyer, W.; Cuomo, C.A.; Wollenburg, K.R.; Bicanic, T.A.; Castañeda, E.; Chang, Y.C.; Chen, J.; et al. The Case for Adopting the "Species Complex" Nomenclature for the Etiologic Agents of Cryptococcosis. mSphere **2017**, 2, e00357-16.
18. Dromer, F.; Gueho, E.; Ronin, O.; Dupont, B. Serotyping of *Cryptococcus neoformans* by using a monoclonal antibody specific for capsular polysaccharide. J Clin Microbiol **1993**, 31, 359–363.
19. Casadevall, A.; Freij, J.B.; Hann-Soden, C.; Taylor, J. Continental Drift and Speciation of the *Cryptococcus neoformans* and *Cryptococcus gattii* Species Complexes. mSphere **2017**, 2, e00103-17.
20. Xu, J.; Vilgalys, R.; Mitchell, T.G. Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*. Mol. Ecol. **2000**, 9, 1471–1481.
21. Ngamskulrungroj, P.; Gilgado, F.; Faganello, J.; Litvintseva, A.P.; Leal, A.L.; Tsui, K.M.; Mitchell, T.G.; Vainstein, M.H.; Meyer, W. Genetic diversity of the *Cryptococcus* species complex suggests that *Cryptococcus gattii* deserves to have varieties. PLoS ONE **2009**, 4, e5862.
22. Sharpton, T.J.; Neafsey, D.E.; Galagan, J.E.; Taylor, J.W. Mechanisms of intron gain and loss in *Cryptococcus*. Genome Biol **2008**, 9, R24.
23. Kwon-Chung, K.J. A New Genus, *Filobasidiella*, the Perfect State of *Cryptococcus neoformans*. Mycologia **1975**, 67, 1197–1200.
24. Kwon-Chung, K.J. A new species of *Filobasidiella*, the sexual state of *Cryptococcus neoformans* B and C serotypes. Mycologia **1976**, 68, 943–946.
25. Lin, X.; Heitman, J. Chlamydospore Formation during Hyphal Growth in *Cryptococcus neoformans*. Eukaryot Cell **2005**, 4, 1746–1754.
26. Lin, X. *Cryptococcus neoformans*: morphogenesis, infection, and evolution. Infect. Genet. Evol. **2009**, 9, 401–416.
27. Lin, X.; Hull, C.M.; Heitman, J. Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*. Nature **2005**, 434, 1017–1021.
28. Zhao, Y.; Lin, J.; Fan, Y.; Lin, X. Life Cycle of *Cryptococcus neoformans*. Annu. Rev. Microbiol. **2019**, 73, 17–42.

29. Hull, C.M.; Heitman, J. Genetics of *Cryptococcus neoformans*. Annu. Rev. Genet. **2002**, 36, 557–615.
30. Wang, P.; Perfect, J.R.; Heitman, J. The G-protein beta subunit GPB1 is required for mating and haploid fruiting in *Cryptococcus neoformans*. Mol. Cell. Biol. **2000**, 20, 352–362.
31. Wickes, B.L.; Mayorga, M.E.; Edman, U.; Edman, J.C. Dimorphism and haploid fruiting in *Cryptococcus neoformans*: association with the alpha-mating type. Proc Natl Acad Sci U S A **1996**, 93, 7327–7331.
32. Lin, X.; Huang, J.C.; Mitchell, T.G.; Heitman, J. Virulence Attributes and Hyphal Growth of *C. neoformans* Are Quantitative Traits and the MATα Allele Enhances Filamentation. PLOS Genetics **2006**, 2, e187.
33. Kwon-Chung, K.J.; Bennett, J.E. Distribution of alpha and alpha mating types of *Cryptococcus neoformans* among natural and clinical isolates. Am. J. Epidemiol. **1978**, 108, 337–340.
34. Yan, Z.; Li, X.; Xu, J. Geographic distribution of mating type alleles of *Cryptococcus neoformans* in four areas of the United States. J. Clin. Microbiol. **2002**, 40, 965–972.
35. Fu, C.; Heitman, J. PRM1 and KAR5 function in cell-cell fusion and karyogamy to drive distinct bisexual and unisexual cycles in the *Cryptococcus* pathogenic species complex. PLoS Genet. **2017**, 13, e1007113.
36. Lin, X.; Heitman, J. The biology of the *Cryptococcus neoformans* species complex. Annu. Rev. Microbiol. **2006**, 60, 69–105.
37. Lin, X.; Litvintseva, A.P.; Nielsen, K.; Patel, S.; Floyd, A.; Mitchell, T.G.; Heitman, J. αADα Hybrids of *Cryptococcus neoformans*: Evidence of Same-Sex Mating in Nature and Hybrid Fitness. PLOS Genetics **2007**, 3, e186.
38. Rhodes, J.; Desjardins, C.A.; Sykes, S.M.; Beale, M.A.; Vanhove, M.; Sakthikumar, S.; Chen, Y.; Gujja, S.; Saif, S.; Chowdhary, A.; et al. Tracing Genetic Exchange and Biogeography of *Cryptococcus neoformans* var. grubii at the Global Population Level. Genetics **2017**, 207, 327–346.
39. Ni, M.; Feretzaki, M.; Li, W.; Floyd-Averette, A.; Mieczkowski, P.; Dietrich, F.S.; Heitman, J. Unisexual and Heterosexual Meiotic Reproduction Generate Aneuploidy and Phenotypic Diversity De Novo in the Yeast *Cryptococcus neoformans*. PLOS Biology **2013**, 11, e1001653.
40. You, M.; Xu, J. The effects of environmental and genetic factors on the germination of basidiospores in the *Cryptococcus gattii* species complex. Sci Rep **2018**, 8, 1–15.

41. Lengeler, K.B.; Cox, G.M.; Heitman, J. Serotype AD strains of *Cryptococcus neoformans* are diploid or aneuploid and are heterozygous at the mating-type locus. Infect. Immun. **2001**, 69, 115–122.
42. Sun, S.; Billmyre, R.B.; Mieczkowski, P.A.; Heitman, J. Unisexual reproduction drives meiotic recombination and phenotypic and karyotypic plasticity in *Cryptococcus neoformans*. PLoS Genet. **2014**, 10, e1004849.
43. Lin, X.; Patel, S.; Litvintseva, A.P.; Floyd, A.; Mitchell, T.G.; Heitman, J. Diploids in the *Cryptococcus neoformans* serotype A population homozygous for the alpha mating type originate via unisexual mating. PLoS Pathog. **2009**, 5, e1000283.
44. Bovers, M.; Hagen, F.; Kuramae, E.E.; Diaz, M.R.; Spanjaard, L.; Dromer, F.; Hoogveld, H.L.; Boekhout, T. Unique hybrids between the fungal pathogens *Cryptococcus neoformans* and *Cryptococcus gattii*. FEMS Yeast Res. **2006**, 6, 599–607.
45. Bovers, M.; Hagen, F.; Kuramae, E.E.; Hoogveld, H.L.; Dromer, F.; St-Germain, G.; Boekhout, T. AIDS Patient Death Caused by Novel *Cryptococcus neoformans* × *C. gattii* Hybrid. Emerg Infect Dis **2008**, 14, 1105–1108.
46. Aminnejad, M.; Diaz, M.; Arabatzis, M.; Castañeda, E.; Lazera, M.; Velegraki, A.; Marriott, D.; Sorrell, T.C.; Meyer, W. Identification of novel hybrids between *Cryptococcus neoformans* var. grubii VNI and *Cryptococcus gattii* VGII. Mycopathologia **2012**, 173, 337–346.
47. Smith, I.M.; Stephan, C.; Hogardt, M.; Klawe, C.; Tintelnot, K.; Rickerts, V. Cryptococcosis due to *Cryptococcus gattii* in Germany from 2004-2013. Int. J. Med. Microbiol. **2015**, 305, 719–723.
48. Hagen, F.; Hare Jensen, R.; Meis, J.F.; Arendrup, M.C. Molecular epidemiology and in vitro antifungal susceptibility testing of 108 clinical *Cryptococcus neoformans* sensu lato and *Cryptococcus gattii* sensu lato isolates from Denmark. Mycoses **2016**, 59, 576–584.
49. Rhodes, J.; Desjardins, C.A.; Sykes, S.M.; Beale, M.A.; Vanhove, M.; Sakthikumar, S.; Chen, Y.; Gujja, S.; Saif, S.; Chowdhary, A.; et al. Population genomics of *Cryptococcus neoformans* var. *grubii* reveals new biogeographic relationships and finely maps hybridization. bioRxiv **2017**, 132894.
50. Bennett, J.E.; Kwon-Chung, K.J.; Howard, D.H. Epidemiologic differences among serotypes of *Cryptococcus neoformans*. Am. J. Epidemiol. **1977**, 105, 582–586.

51. Maduro, A.P.; Mansinho, K.; Teles, F.; Silva, I.; Meyer, W.; Martins, M.L.; Inácio, J. Insights on the genotype distribution among *Cryptococcus neoformans* and *C. gattii* Portuguese clinical isolates. Curr. Microbiol. **2014**, 68, 199–203.
52. Cogliati, M.; Esposto, M.C.; Clarke, D.L.; Wickes, B.L.; Viviani, M.A. Origin of *Cryptococcus neoformans* var. *neoformans* diploid strains. J. Clin. Microbiol. **2001**, 39, 3889–3894.
53. Viviani, M.A.; Antinori, S.; Cogliati, M.; Esposto, M.C.; Pinsi, G.; Casari, S.; Bergamasco, A.F.; Santis, M.D.; Ghirga, P.; Bonaccorso, C.; et al. European Confederation of Medical Mycology (ECMM) prospective survey of cryptococcosis: report from Italy. Medical mycology **2002**, 40, 507–517.
54. Samarasinghe, H.; Xu, J. Hybrids and hybridization in the *Cryptococcus neoformans* and *Cryptococcus gattii* species complexes. Infect. Genet. Evol. **2018**, 66, 245–255.
55. Sun, S.; Xu, J. Genetic analyses of a hybrid cross between serotypes A and D strains of the human pathogenic fungus *Cryptococcus neoformans*. Genetics **2007**, 177, 1475–1486.
56. Li, W.; Averette, A.F.; Desnos-Ollivier, M.; Ni, M.; Dromer, F.; Heitman, J. Genetic Diversity and Genomic Plasticity of *Cryptococcus neoformans* AD Hybrid Strains. G3 (Bethesda) **2012**, 2, 83–97.
57. Campbell, L.T.; Fraser, J.A.; Nichols, C.B.; Dietrich, F.S.; Carter, D.; Heitman, J. Clinical and Environmental Isolates of *Cryptococcus gattii* from Australia That Retain Sexual Fecundity. Eukaryotic Cell **2005**, 4, 1410–1419.
58. Kidd, S.E.; Hagen, F.; Tscharke, R.L.; Huynh, M.; Bartlett, K.H.; Fyfe, M.; Macdougall, L.; Boekhout, T.; Kwon-Chung, K.J.; Meyer, W. A rare genotype of *Cryptococcus gattii* caused the cryptococcosis outbreak on Vancouver Island (British Columbia, Canada). Proc. Natl. Acad. Sci. U.S.A. **2004**, 101, 17258–17263.
59. Fraser, J.A.; Subaran, R.L.; Nichols, C.B.; Heitman, J. Recapitulation of the sexual cycle of the primary fungal pathogen *Cryptococcus neoformans* var. *gattii*: implications for an outbreak on Vancouver Island, Canada. Eukaryotic Cell **2003**, 2, 1036–1045.
60. Voelz, K.; Ma, H.; Phadke, S.; Byrnes, E.J.; Zhu, P.; Mueller, O.; Farrer, R.A.; Henk, D.A.; Lewit, Y.; Hsueh, Y.-P.; et al. Transmission of Hypervirulence Traits via Sexual Reproduction within and between Lineages of the Human Fungal Pathogen *Cryptococcus gattii*. PLOS Genetics **2013**, 9, e1003771.
61. Vogan, A.A.; Khankhet, J.; Xu, J. Evidence for mitotic recombination within the basidia of a hybrid cross of *Cryptococcus neoformans*. PLoS ONE **2013**, 8, e62790.

62. Forsythe, A.; Vogan, A.; Xu, J. Genetic and environmental influences on the germination of basidiospores in the *Cryptococcus neoformans* species complex. Sci Rep **2016**, 6, 1–12.
63. Tanaka, R.; Nishimura, K.; Miyaji, M. Ploidy of serotype AD strains of *Cryptococcus neoformans*. Nihon Ishinkin Gakkai Zasshi **1999**, 40, 31–34.
64. Desnos-Ollivier, M.; Patel, S.; Raoux-Barbot, D.; Heitman, J.; Dromer, F.; French Cryptococcosis Study Group Cryptococcosis Serotypes Impact Outcome and Provide Evidence of *Cryptococcus neoformans* Speciation. MBio **2015**, 6, e00311.
65. Litvintseva, A.P.; Lin, X.; Templeton, I.; Heitman, J.; Mitchell, T.G. Many Globally Isolated AD Hybrid Strains of *Cryptococcus neoformans* Originated in Africa. PLOS Pathogens **2007**, 3, e114.
66. Lin, X.; Nielsen, K.; Patel, S.; Heitman, J. Impact of Mating Type, Serotype, and Ploidy on the Virulence of *Cryptococcus neoformans*. Infect Immun **2008**, 76, 2923–2938.
67. Li, M.; Liao, Y.; Chen, M.; Pan, W.; Weng, L. Antifungal susceptibilities of *Cryptococcus* species complex isolates from AIDS and non-AIDS patients in Southeast China. Braz J Infect Dis **2012**, 16, 175–179.
68. Vogan, A.A.; Khankhet, J.; Samarasinghe, H.; Xu, J. Identification of QTLs Associated with Virulence Related Traits and Drug Resistance in *Cryptococcus neoformans*. G3 (Bethesda) **2016**, 6, 2745–2759.
69. Sun, S.; Xu, J. Chromosomal rearrangements between serotype A and D strains in *Cryptococcus neoformans*. PLoS ONE **2009**, 4, e5524.
70. Hu, G.; Liu, I.; Sham, A.; Stajich, J.E.; Dietrich, F.S.; Kronstad, J.W. Comparative hybridization reveals extensive genome variation in the AIDS-associated pathogen *Cryptococcus neoformans*. Genome Biol **2008**, 9, R41.
71. Kavanaugh, L.A.; Fraser, J.A.; Dietrich, F.S. Recent evolution of the human pathogen *Cryptococcus neoformans* by intervarietal transfer of a 14-gene fragment. Mol. Biol. Evol. **2006**, 23, 1879–1890.
72. Vogan, A.A.; Xu, J. Evidence for genetic incompatibilities associated with post-zygotic reproductive isolation in the human fungal pathogen *Cryptococcus neoformans*. Genome **2014**, 57, 335–344.
73. Cogliati, M.; Esposto, M.C.; Tortorano, A.M.; Viviani, M.A. *Cryptococcus neoformans* population includes hybrid strains homozygous at mating-type locus. FEMS Yeast Res. **2006**, 6, 608–613.

74. Llorente, B.; Smith, C.E.; Symington, L.S. Break-induced replication: what is it and what is it for? Cell Cycle **2008**, 7, 859–864.
75. Bennett, R.J.; Forche, A.; Berman, J. Rapid mechanisms for generating genome diversity: whole ploidy shifts, aneuploidy, and loss of heterozygosity. Cold Spring Harb Perspect Med **2014**, 4, a019604.
76. Samarasinghe, H.; Vogan, A.; Pum, N.; Xu, J. Patterns of allele distribution in a hybrid population of the *Cryptococcus neoformans* species complex. Mycoses **2019**.
77. Dong, K.; You, M.; Xu, J. Genetic changes in experimental populations of a hybrid in the *Cryptococcus neoformans* species complex. Pathogens **2020**, 9, 3.
78. Sionov, E.; Lee, H.; Chang, Y.C.; Kwon-Chung, K.J. *Cryptococcus neoformans* overcomes stress of azole drugs by formation of disomy in specific multiple chromosomes. PLoS Pathog. **2010**, 6, e1000848.
79. Hata, K.; Ohkusu, M.; Aoki, S.; Ito-Kuwa, S.; Pienthaweechai, K.; Takeo, K. Cells of different ploidy are often present together in *Cryptococcus neoformans* strains. Nihon Ishinkin Gakkai Zasshi **2000**, 41, 161–167.
80. Toffaletti, D.L.; Nielsen, K.; Dietrich, F.; Heitman, J.; Perfect, J.R. *Cryptococcus neoformans* mitochondrial genomes from serotype A and D strains do not influence virulence. Curr. Genet. **2004**, 46, 193–204.
81. Skosireva, I.; James, T.Y.; Sun, S.; Xu, J. Mitochondrial inheritance in haploid x non-haploid crosses in *Cryptococcus neoformans*. Curr. Genet. **2010**, 56, 163–176.
82. Zaragoza, O.; Nielsen, K. Titan cells in *Cryptococcus neoformans*: cells with a giant impact. Curr. Opin. Microbiol. **2013**, 16, 409–413.
83. Hu, G.; Wang, J.; Choi, J.; Jung, W.H.; Liu, I.; Litvintseva, A.P.; Bicanic, T.; Aurora, R.; Mitchell, T.G.; Perfect, J.R.; et al. Variation in chromosome copy number influences the virulence of *Cryptococcus neoformans* and occurs in isolates from AIDS patients. BMC Genomics **2011**, 12, 526.
84. Fan, Y.; Lin, X. Multiple Applications of a Transient CRISPR-Cas9 Coupled with Electroporation (TRACE) System in the *Cryptococcus neoformans* Species Complex. Genetics **2018**, 208, 1357–1372.
85. Xie, G.Y.; Olson, D.H.; Blaustein, A.R. Projecting the Global Distribution of the Emerging Amphibian Fungal Pathogen, Batrachochytrium dendrobatidis, Based on IPCC Climate Futures. PLoS One **2016**, 11, e0160746.
86. Berger, L.; Speare, R.; Daszak, P.; Green, D.E.; Cunningham, A.A.; Goggin, C.L.; Slocombe, R.; Ragan, M.A.; Hyatt, A.D.; McDonald, K.R.; et al. Chytridiomycosis

causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. PNAS **1998**, 95, 9031–9036.

87. Rachowicz, L.J.; Knapp, R.A.; Morgan, J.A.T.; Stice, M.J.; Vredenburg, V.T.; Parker, J.M.; Briggs, C.J. Emerging Infectious Disease as a Proximate Cause of Amphibian Mass Mortality. Ecology **2006**, 87, 1671–1683.
88. Olson, D.H.; Aanensen, D.M.; Ronnenberg, K.L.; Powell, C.I.; Walker, S.F.; Bielby, J.; Garner, T.W.J.; Weaver, G.; Bd Mapping Group; Fisher, M.C. Mapping the global emergence of *Batrachochytrium dendrobatidis*, the amphibian chytrid fungus. PLoS ONE **2013**, 8, e56802.
89. O'Hanlon, S.J.; Rieux, A.; Farrer, R.A.; Rosa, G.M.; Waldman, B.; Bataille, A.; Kosch, T.A.; Murray, K.A.; Brankovics, B.; Fumagalli, M.; et al. Recent Asian origin of chytrid fungi causing global amphibian declines. Science **2018**, 360, 621–627.
90. Farrer, R.A.; Weinert, L.A.; Bielby, J.; Garner, T.W.J.; Balloux, F.; Clare, F.; Bosch, J.; Cunningham, A.A.; Weldon, C.; Preez, L.H. du; et al. Multiple emergences of genetically diverse amphibian-infecting chytrids include a globalized hypervirulent recombinant lineage. PNAS **2011**, 108, 18732–18736.
91. Schloegel, L.M.; Toledo, L.F.; Longcore, J.E.; Greenspan, S.E.; Vieira, C.A.; Lee, M.; Zhao, S.; Wangen, C.; Ferreira, C.M.; Hipolito, M.; et al. Novel, panzootic and hybrid genotypes of amphibian chytridiomycosis associated with the bullfrog trade. Mol. Ecol. **2012**, 21, 5162–5177.
92. Byrne, A.Q.; Vredenburg, V.T.; Martel, A.; Pasmans, F.; Bell, R.C.; Blackburn, D.C.; Bletz, M.C.; Bosch, J.; Briggs, C.J.; Brown, R.M.; et al. Cryptic diversity of a widespread global pathogen reveals expanded threats to amphibian conservation. Proc. Natl. Acad. Sci. U.S.A. **2019**, 116, 20382–20387.
93. James, T.Y.; Litvintseva, A.P.; Vilgalys, R.; Morgan, J.A.T.; Taylor, J.W.; Fisher, M.C.; Berger, L.; Weldon, C.; du Preez, L.; Longcore, J.E. Rapid global expansion of the fungal disease chytridiomycosis into declining and healthy amphibian populations. PLoS Pathog. **2009**, 5, e1000458.
94. Rosenblum, E.B.; James, T.Y.; Zamudio, K.R.; Poorten, T.J.; Ilut, D.; Rodriguez, D.; Eastman, J.M.; Richards-Hrdlicka, K.; Joneson, S.; Jenkinson, T.S.; et al. Complex history of the amphibian-killing chytrid fungus revealed with genome resequencing data. Proc. Natl. Acad. Sci. U.S.A. **2013**, 110, 9385–9390.
95. Jenkinson, T.S.; Betancourt Román, C.M.; Lambertini, C.; Valencia-Aguilar, A.; Rodriguez, D.; Nunes-de-Almeida, C.H.L.; Ruggeri, J.; Belasen, A.M.; da Silva Leite,

D.; Zamudio, K.R.; et al. Amphibian-killing chytrid in Brazil comprises both locally endemic and globally expanding populations. Mol. Ecol. **2016**, 25, 2978–2996.

96. Giraud, T.; Enjalbert, J.; Fournier, E.; Delmote, F.; Dutech, C. Population genetics of fungal diseases of plants. Parasite **2008**, 15, 449–454.
97. Bennett, R.J. The parasexual lifestyle of Candida albicans. Curr. Opin. Microbiol. **2015**, 28, 10–17.
98. Greenspan, S.E.; Lambertini, C.; Carvalho, T.; James, T.Y.; Toledo, L.F.; Haddad, C.F.B.; Becker, C.G. Hybrids of amphibian chytrid show high virulence in native hosts. Sci Rep **2018**, 8, 1–10.
99. Becker, C.G.; Rodriguez, D.; Toledo, L.F.; Longo, A.V.; Lambertini, C.; Corrêa, D.T.; Leite, D.S.; Haddad, C.F.B.; Zamudio, K.R. Partitioning the net effect of host diversity on an emerging amphibian pathogen. Proc. Biol. Sci. **2014**, 281, 20141796.
100. Farrer, R.A.; Henk, D.A.; Garner, T.W.J.; Balloux, F.; Woodhams, D.C.; Fisher, M.C. Chromosomal Copy Number Variation, Selection and Uneven Rates of Recombination Reveal Cryptic Genome Diversity Linked to Pathogenicity. PLOS Genetics **2013**, 9, e1003703.
101. Daum, J.M.; Davis, L.R.; Bigler, L.; Woodhams, D.C. Hybrid advantage in skin peptide immune defenses of water frogs (*Pelophylax esculentus*) at risk from emerging pathogens. Infection, Genetics and Evolution **2012**, 12, 1854–1864.
102. Refsnider, J.M.; Poorten, T.J.; Langhammer, P.F.; Burrowes, P.A.; Rosenblum, E.B. Genomic Correlates of Virulence Attenuation in the Deadly Amphibian Chytrid Fungus, *Batrachochytrium dendrobatidis*. G3 (Bethesda) **2015**, 5, 2291–2298.
103. Langhammer, P.F.; Lips, K.R.; Burrowes, P.A.; Tunstall, T.; Palmer, C.M.; Collins, J.P. A fungal pathogen of amphibians, *Batrachochytrium dendrobatidis*, attenuates in pathogenicity with in vitro passages. PLoS ONE **2013**, 8, e77630.
104. James, T.Y.; Toledo, L.F.; Rödder, D.; da Silva Leite, D.; Belasen, A.M.; Betancourt-Román, C.M.; Jenkinson, T.S.; Soto-Azat, C.; Lambertini, C.; Longo, A.V.; et al. Disentangling host, pathogen, and environmental determinants of a recently emerged wildlife disease: lessons from the first 15 years of amphibian chytridiomycosis research. Ecol Evol **2015**, 5, 4079–4097.
105. Joneson, S.; Stajich, J.E.; Shiu, S.-H.; Rosenblum, E.B. Genomic transition to pathogenicity in chytrid fungi. PLoS Pathog. **2011**, 7, e1002338.
106. Rosenblum, E.B.; Stajich, J.E.; Maddox, N.; Eisen, M.B. Global gene expression profiles for life stages of the deadly amphibian pathogen *Batrachochytrium dendrobatidis*. Proc. Natl. Acad. Sci. U.S.A. **2008**, 105, 17034–17039.

107. Rosenblum, E.B.; Poorten, T.J.; Joneson, S.; Settles, M. Substrate-Specific Gene Expression in Batrachochytrium dendrobatidis, the Chytrid Pathogen of Amphibians. PLOS ONE **2012**, 7, e49924.
108. Saikkonen, K.; Young, C.A.; Helander, M.; Schardl, C.L. Endophytic Epichloë species and their grass hosts: from evolution to applications. Plant Mol. Biol. **2016**, 90, 665–675.
109. Inderbitzin, P.; Davis, R.M.; Bostock, R.M.; Subbarao, K.V. The ascomycete *Verticillium longisporum* is a hybrid and a plant pathogen with an expanded host range. PLoS ONE **2011**, 6, e18260.
110. Fogelqvist, J.; Tzelepis, G.; Bejai, S.; Ilbäck, J.; Schwelm, A.; Dixelius, C. Analysis of the hybrid genomes of two field isolates of the soil-borne fungal species Verticillium longisporum. BMC Genomics **2018**, 19, 14.
111. Marcet-Houben, M.; Gabaldón, T. Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. PLOS Biology **2015**, 13, e1002220.
112. Fay, J.C.; Liu, P.; Ong, G.T.; Dunham, M.J.; Cromie, G.A.; Jeffery, E.W.; Ludlow, C.L.; Dudley, A.M. A polyploid admixed origin of beer yeasts derived from European and Asian wine populations. PLoS Biol. **2019**, 17, e3000147.
113. Langdon, Q.K.; Peris, D.; Baker, E.P.; Opulente, D.A.; Nguyen, H.-V.; Bond, U.; Gonçalves, P.; Sampaio, J.P.; Libkind, D.; Hittinger, C.T. Fermentation innovation through complex hybridization of wild and domesticated yeasts. Nat Ecol Evol **2019**, 3, 1576–1586.
114. Gostinčar, C.; Stajich, J.E.; Zupančič, J.; Zalar, P.; Gunde-Cimerman, N. Genomic evidence for intraspecific hybridization in a clonal and extremely halotolerant yeast. BMC Genomics **2018**, 19, 364.
115. Pryszcz, L.P.; Németh, T.; Saus, E.; Ksiezopolska, E.; Hegedűsová, E.; Nosek, J.; Wolfe, K.H.; Gacser, A.; Gabaldón, T. The Genomic Aftermath of Hybridization in the Opportunistic Pathogen *Candida metapsilosis*. PLoS Genet. **2015**, 11, e1005626.
116. Stukenbrock, E.H.; Christiansen, F.B.; Hansen, T.T.; Dutheil, J.Y.; Schierup, M.H. Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. PNAS **2012**, 109, 10954–10959.
117. Le Gac, M.; Hood, M.E.; Fournier, E.; Giraud, T. Phylogenetic evidence of host-specific cryptic species in the anther smut fungus. Evolution **2007**, 61, 15–26.

118. Turner, E.; Jacobson, D.J.; Taylor, J.W. Genetic Architecture of a Reinforced, Postmating, Reproductive Isolation Barrier between Neurospora Species Indicates Evolution via Natural Selection. PLOS Genetics **2011**, 7, e1002204.
119. Büker, B.; Petit, E.; Begerow, D.; Hood, M.E. Experimental hybridization and backcrossing reveal forces of reproductive isolation in Microbotryum. BMC Evol. Biol. **2013**, 13, 224.
120. Olson, A.; Stenlid, J. Plant pathogens. Mitochondrial control of fungal hybrid virulence. Nature **2001**, 411, 438.

Chapter 7

Patterns of allele distribution in a hybrid population of the *Cryptococcus neoformans* species complex

7.1 Preface

The increasing prevalence of *C. neoformans* x *C. deneoformans* hybrids (termed AD hybrids) in clinical settings warrants investigations into their genetic makeup and potential for enhanced virulence. Previous studies have shown that, unlike their haploid parents, AD hybrids are diploid or aneuploid with significant genomic instability. We used polymerase chain reaction – restriction fragment length polymorphism (PCR-RFLP) markers to characterize the genome-wide allele distribution of a lab-generated AD population. Our results showed AD hybrids to have significantly skewed allele distributions with alleles from the mitochondria-donor parent being significantly favored over the other. Our study contributes an important piece to painting a comprehensive picture of the genetic architecture of these hybrids.

This manuscript has been published in Mycoses volume 63, issue 3. I am the primary contributor of this work. I conducted the majority of the experiments, data analysis and drafting of the manuscript. Co-authors of this article are Aaron Vogan, Nicole Pum and Jianping Xu.

7.2 Abstract

Background: The sister yeast species *Cryptococcus neoformans* (serotype A) and *Cryptococcus deneoformans* (serotype D) are causative agents of deadly cryptococcosis and fungal meningoencephalitis. These haploid yeasts can hybridize in nature, giving rise to AD hybrids that are predominantly diploid or aneuploid. Despite their increasing prevalence in clinical settings, much remains unknown about the allelic distribution patterns in AD hybrid strains.

Objectives: This study aims to characterize allele distributions in AD hybrids derived from the same basidium as well as from multiple basidia in a laboratory *C. neoformans* x *C. deneoformans* hybrid cross.

Methods: We dissected a total of 1625 basidiospores from 31 basidia. The 297 basidiospores that successfully germinated were genotyped by molecular characterization of 33 markers using PCR-RFLP, with at least two markers on each of the 14 chromosomes in the genome.

Results: 294 of the 297 strains contained at least one heterozygous locus, with a mean heterozygosity of ~30% per strain. Most hybrid genomes and chromosomes displayed significantly distorted allele distributions, with offspring originating from the same basidium tended to have alleles at different loci from the same parent. More basidia were skewed in favor of *C. deneoformans* alleles, the mitochondria-donor parent, than the *C. neoformans* alleles.

Conclusions: The divergence between *C. neoformans* and *C. deneoformans* genomes has likely created co-adapted allelic combinations, with their co-segregation in hybrid offspring imparting a significant fitness benefit. However, the diversity of genotypes recovered here in a single hybridization event indicates the enormous capacity of AD hybrids for adaptation and diversification.

7.3 Introduction

Hybridization between closely related lineages is a common occurrence in nature among plants, animals and eukaryotic microbes such as fungi. Depending on the compatibility between the parental lineages, outcomes of hybridization can range from complete prezygotic isolation to production of fertile progeny with hybrid vigor. The study of hybrids can provide valuable insights into the evolutionary history, genetic compatibilities, and phenotypic profiles of the parents. Existence of hybrids can also complicate nomenclature assignments and render taxonomic statuses of the parental species ambiguous. The pathogenic yeast pair *Cryptococcus neoformans* (serotype A) and *Cryptococcus deneoformans* (serotype D) of the *C. neoformans* species complex (CNSC), known to hybridize in nature, provides a valuable model system for studying these phenomena.

C. neoformans and *C. deneoformans* are the causative agents of cryptococcosis that claims over 180 000 lives worldwide every year[1]. Commonly found in the environment in association with trees, soil and bird droppings[2], their combined geographical distribution spans the entire globe with isolates having been recovered from all continents except Antarctica[3]. Haploid *C. neoformans* and *C. deneoformans* strains can hybridize with each other in nature to produce diploid/aneuploid hybrid strains, commonly referred to as AD hybrids[4,5]. Initially discovered in the late 1970s, AD hybrids have since been recovered from both environmental and clinical samples with a high prevalence in Mediterranean Europe, where up to 30% of all cryptococcal infections are caused by hybrid strains[6–10]. AD hybrids can also be generated under laboratory conditions by mating strains of the two species with each other, albeit with low spore germination rates of ~5-20%[5,11,12]. Hybrid spores that do germinate often contain an abnormal number of chromosomes, consistent with impaired meiosis and nondisjunction of parental chromosomes during hybridization[4,5,13]. While most AD hybrids are heterozygous at multiple loci across the genome, including the mating type (*MAT*) locus, few are self-fertile or can mate with strains of the opposite mating type[5,14].

Despite low viability and infertility of sexual spores derived from hybrid crosses, both lab-derived and natural AD hybrids show a remarkable phenotypic diversity in

virulence and associated characteristics. While some studies found AD hybrids to show hybrid vigor, others have reported the virulence of AD hybrids to be comparable to that of their more virulent parent[15–17]. Both the hybrids' genetic background (including mating type) and the type of host animals used as infection models can influence the *in vivo* virulence variations[15,17]. Furthermore, a few studies found AD hybrids to exceed both parents in their resistance to UV irradiation and the antifungal drug FK506[18–22]. Another study found extreme phenotypes, both positive (i.e. significantly more than both parents) and negative (i.e. significantly less than both parents), among lab-derived AD hybrids in the expression of virulence factors including capsule production, growth at 37°C and melanin production[16]. The phenotypic diversity found among AD hybrids in clinical, environmental and laboratory settings is indicative of significant genetic diversity present within the hybrid population, a likely product of their novel and unique genetic make-up.

Hybridization between *C. neoformans* and *C. deneoformans* is marked with frequent chromosomal non-disjunction, meiotic recombination and mitotic recombination, leading to the creation of novel allelic combinations[11]. Furthermore, AD hybrid genomes can be subjected to loss of heterozygosity (LOH) during vegetative growth via chromosomal loss and/or mitotic gene conversion[22,23]. While whole-genome hybridization and sequencing approaches have been used to study cryptococcal hybrid genetics[24,25], most such studies have been limited by a small sample size. As a result, much remains unknown about allelic distributions and compatibilities at the chromosomal level, especially for AD hybrids derived from the same hybridization event. Studying allele segregation in AD hybrid genomes will provide much needed insights into their genetic repertoire and phenotypic diversity.

A topic currently under debate is whether *C. neoformans* and *C. deneoformans*, estimated to have diverged from a common ancestor ~20 million years ago[26,27], should be classified as varieties, as opposed to two distinct species[28–30]. At present, *C. neoformans* and *C. deneoformans* are alternatively referred to as *C. neoformans* var. *grubii* and *C. neoformans* var. *neoformans* respectively. AD hybrids have complicated this issue of classification as their existence attests to the ability of the two parental species to mate with each other while low viability of hybrids points to some post-

zygotic reproductive barriers. Studying the genetic composition of viable AD hybrids can shed light on the structures and compatibilities of *C. neoformans* and *C. deneoformans* genomes, which in turn will provide insights into the extent of genotypic and reproductive divergence between the two lineages.

The objective of this study was to investigate the patterns of genome-wide allele distribution in a lab-derived AD hybrid population. We were specifically interested in characterizing allele segregations at three distinct levels: (i) general trends in genome-wide allele distribution at the population level; (ii) differential distribution patterns between chromosomes; and (iii) distributions among spores originating from the same basidium (i.e. same meiotic event). To accomplish these objectives, we constructed a hybrid cross in the lab by mating CDC15 (*C. neoformans*, serotype A, *MAT***α**) with JEC20 (*C. deneoformans*, serotype D, *MAT***a**). We dissected a total of 1625 basidiospores from 31 basidia: the positions of dissected spores on the spore chains were recorded to follow the sequential rounds of mitosis that occur during spore production within each basidium. 297 spores that successfully germinated were then genotyped at 33 PCR-RFLP (Polymerase Chain Reaction – Restriction Fragment Length Polymorphism) markers dispersed throughout the genome and their patterns of allelic distributions were analyzed.

7.4 Materials and Methods

Hybrid population: *C. neoformans* strain CDC15 (serotype A, molecular type VNI, *MAT***α**) was mated with *C. deneoformans* strain JEC20 (serotype D, molecular type VNIV, *MAT***a**) on V8 juice agar medium for ~4 weeks until a prominent ring of sexual mycelia was observed around the mating colonies. Basidiospores were dissected from the spore chains onto yeast extract peptone dextrose (YEPD) agar using a Singer MSM 300 micromanipulator as described in Vogan et al. (2013)[11]. Spores were dissected starting from the ends of the chains and then sequentially moving towards the basidia. Spores closer to the basidia contain nuclei that have undergone more mitotic divisions than those located further away from the basidia. Basidiospores were allowed to germinate over 2 weeks at room temperature. The germination rate per basidium was calculated as the number of spores that germinated out of the total number of spores that were micro-dissected from that basidium onto YEPD.

Genotyping: Genomic DNA was extracted from spores that germinated to form colonies on YEPD using standard chloroform-isoamyl alcohol methodology[11,26]. Hybrid strains were then genotyped at 33 co-dominant PCR-RFLP markers (**Supplementary Table S1**) taken from a previous study[16]. At least two markers were selected for each of the 14 chromosomes in the genome (**Figure 1**). The primer sequences and restriction enzymes of the PCR-RFLP markers used in the study are listed in **Supplementary Table S1**. At each marker, the genotypes were recorded as 'A' for the *C. neoformans* (serotype A) allele, 'D' for the *C. deneoformans* (serotype D) allele, and 'AD' for the presence of both serotype A and serotype D alleles.

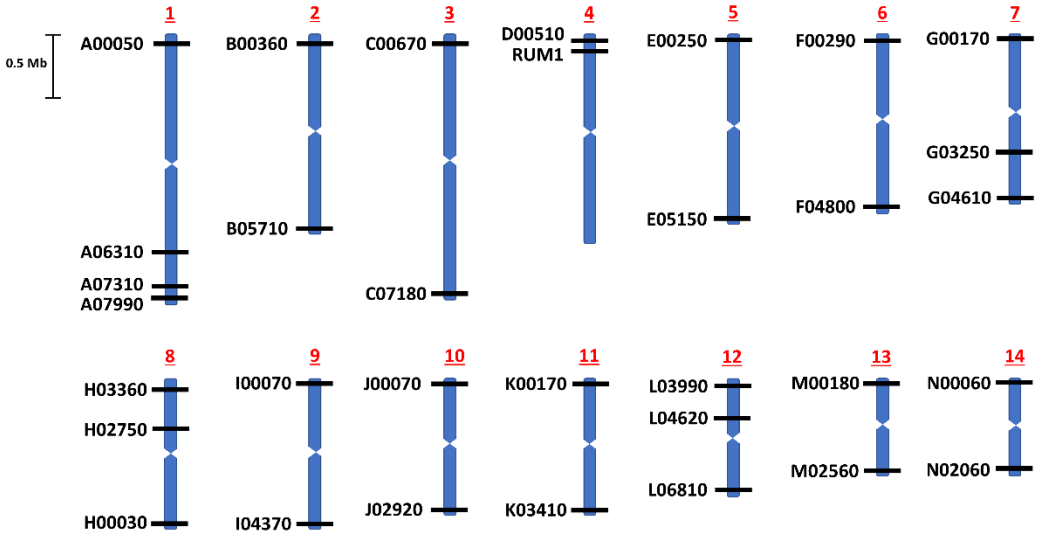


Figure 1: The positions of the 33 PCR-RFLP markers on chromosomes. The AD hybrids were genotyped at each of these 33 PCR-RFLP markers to determine whether they contained one or both parental alleles at these loci. Most markers are located at the ends of chromosomes with a few located in the middle.

Trends in Heterozygosity: The percent heterozygosity for each progeny strain was calculated as the proportion of loci that contained both A and D alleles, out of the total of 33 genotyped loci. Then, for each basidium, the progeny strains were listed in the order they appeared on the spore chain and their numbers of heterozygous loci were compared to determine if there was a gradual change of heterozygosity along each spore chain.

Distorted allele segregations: For each progeny strain, the numbers of loci that were homo-/hemi-zygous for the A allele, homo-/hemi-zygous for the D allele, and

heterozygous for both A and D alleles were counted. Allele distributions were considered to be skewed if a hybrid strain contained significantly more loci that were homo-/hemi-zygous for one parental allele over the other. Next, genotype data for progeny strains originating from the same basidium were combined to detect any common patterns in allele distributions for marker loci across entire basidia.

Preferential retention of chromosomes from one parent: The genotype data from all hybrid strains for markers located on the same chromosome were pooled together. Then, the proportions (p) of progeny strains that were heterozygous (AD) or homo-/hemi-zygous for the parental alleles (either AA or DD) were calculated for each of the 14 chromosomes. Next, observed frequencies (f) of the two parental alleles at each chromosome were calculated assuming diploidy at all loci. Equation 1 shows a sample calculation for determining the observed frequency of the A allele at a chromosome. The expected frequency for each of the two parental alleles was 0.5 under the assumption of unbiased, random segregation.

Equation 1

$$f(A) = \frac{2p(AA) + p(AD)}{200}$$

Chi-square tests were performed for each chromosome to test whether observed frequencies were significantly different from expected frequencies at a significance threshold of $p < 0.05$.

Bateson–Dobzhansky–Muller (BDM) incompatibilities: The BDM incompatibility model explains a form of post-zygotic reproductive isolation where alleles at two different loci of two divergent populations are no longer compatible with each other[31,32]. The current study included two PCR-RFLP markers (A6310 and A7310) that are ~280 kb apart on chromosome 1 between which a unidirectional BDM incompatibility was detected in a previous study[33]. In that case, the *C. neoformans* allele at A6310 was found to be incompatible with the *C. deneoformans* allele at A7310. The genotypes of our AD hybrid population were analyzed for the presence of this incompatibility.

Ethics statement: The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to and the appropriate ethical review committee approval has been received. This study did not involve any human nor animal subjects.

7.5 Results

Spore germination: A total of 1625 spores from 31 basidia were micro-dissected from the periphery of the mating colonies between CDC15 and JEC20. The number of basidiospores dissected per basidium ranged from 25 to 71 (**Supplementary Table S2**). For 11 of the 31 basidia, none of the dissected basidiospores germinated and those spores were assumed to be inviable under our germination condition. 297 basidiospores from 20 basidia successfully germinated to produce colonies on YEPD agar, resulting in an overall germination rate of 18.3%. Among the 20 basidia that contained viable spores, germination rates varied significantly from 7.1% in basidium 7 to 73.5% in basidium 3, with an average germination rate of 27.1% (**Supplementary Table S2**). Germination rates of basidia were not significantly correlated with the total number of dissected basidiospores ($p > 0.3$). The position of the germinated spores on their respective spore chains are shown in **Supplementary Table S2.** We would like to note that for most spores, the position on the chain is given as a range as it was difficult to determine the exact position due to their small size (~3-5µM) and the clustered nature of the spore chains[34].

Heterozygosity: We successfully genotyped the 297 hybrid strains at the 33 PCR-RFLP markers. 294 strains contained at least one heterozygous locus while the remaining three were homo/hemizygous for A or D alleles across all 33 loci (**Figure 2**). None of the hybrid strains were heterozygous at all 33 loci. The average heterozygosity per strain in the population was 30% (~10 out of 33 loci). Significant variability in heterozygosity was observed among hybrid strains, with a range of 0% to 70% (**Figures 2 and 3**). Interestingly, strains derived from the same basidium tended to have similar levels of heterozygosity. For example, the average heterozygosity across the nine progeny strains of basidium 2 was 5%, well below the average of the total population. On the other hand, the 28 progeny strains of basidium

9 displayed higher heterozygosity (45.8%) than the population average.

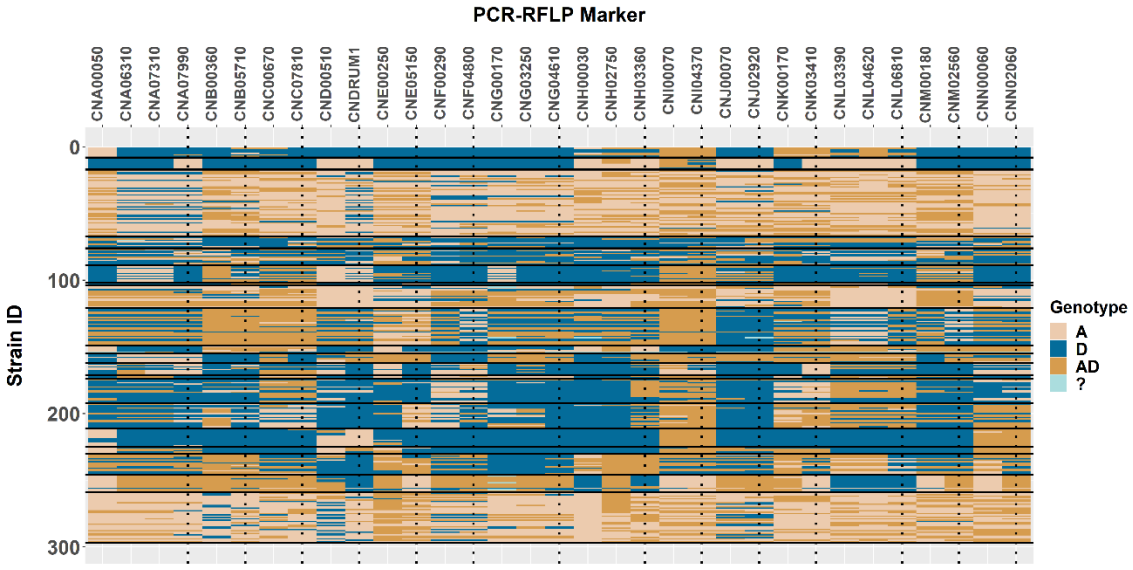


Figure 2: Genotypes of AD hybrids at the 33 PCR-RFLP markers. Each row represents a hybrid strain and each column is a separate marker. The boxes are coloured according to the genotypes—A, D, AD and unknown. Strains originating from the same basidium are grouped together by the black, horizontal lines. The black, vertical lines group markers located on the same chromosome.

Limited evidence for loss of heterozygosity during basidiospore formation: Within each basidium, we ordered the progeny strains according to their positions on the spore chains and compared their numbers of heterozygous loci. In basidium 2, the first 1-3 spores on the spore chain contained two or three heterozygous loci whereas the last two spores that germinated (7^{th} and 8^{th} on the spore chain) were homo/hemizygous at all 33 loci (**Supplementary Table S2** and **Figure 3**), consistent with gradual loss of heterozygosity as more basidiospores were produced. However, apart from this basidium, we failed to see any conclusive evidence for decreasing heterozygosity with an increasing number of mitotic divisions within basidia. The number of heterozygous loci often fluctuated between strains irrespective of their positions on the spore chains (**Figure 3**).

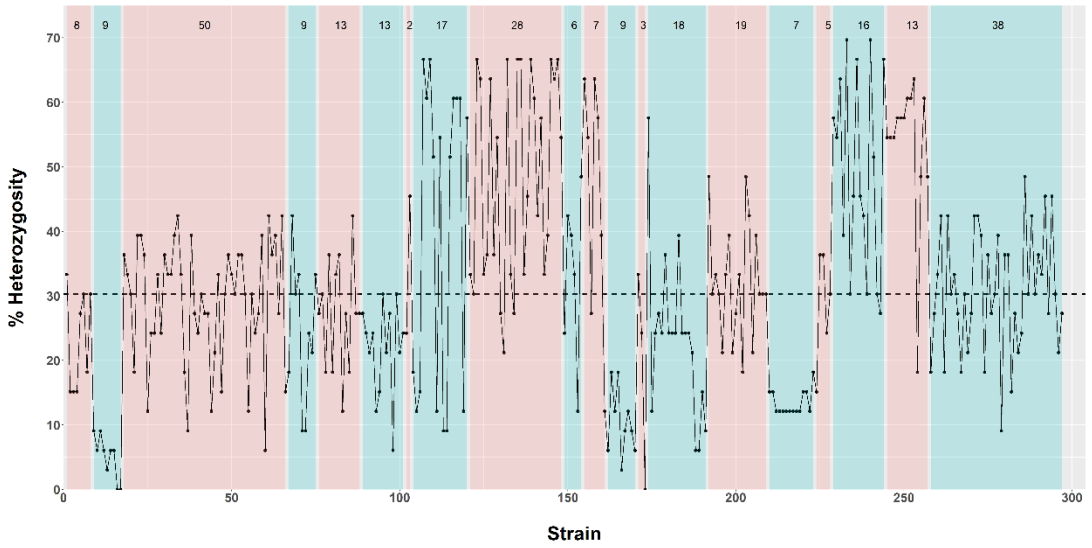


Figure 3: Heterozygosity of hybrid strains. The per cent of heterozygosity (number of heterozygous loci/a total of 33 genotyped loci) of the 297 AD hybrid strains is shown here. Strains are grouped together according to their originating basidium, with different basidia visualised as different-coloured blocks. The number of progeny strains belonging to each basidium is shown at the top of the plot.

Genetic heterogeneity among hybrids: We concatenated the genotypes at the 33 loci to obtain a multilocus genotype (MLG) for each hybrid progeny. Most hybrid strains originating from the same basidium showed unique MLGs. For example, each of the three progeny strains derived from basidium 13 had a unique genotype, while the 50 progeny strains derived from basidium 3 belonged to 46 different MLGs (**Figure 2 and Supplementary Table S2**). The number of unique MLGs present within a basidium was strongly correlated with its germination rate ($r = 0.874$, $p < 0.001$).

Genome-wide bias in allele distributions: In most hybrid strains, allele distributions across the genome were skewed in favor of one parent over the other. Here we defined allele skewness in a progeny strain as the number of loci (out of the 33) homo-/hemi-zygous for one parental allele being greater than twice (≥ x2) the number of loci that was homo/hemizygous for the other parental allele. The most extreme case was progeny strain 98 in which 30 of the 33 genotyped loci were homo-/hemi-zygous for the serotype D allele: of the three remaining loci, two were heterozygous and the third contained the serotype A allele. Overall, 146 out of the 297 strains were skewed for *C. deneoformans* alleles (49%), 106/297 were skewed for *C. neoformans* alleles (36%) and the remaining 45 strains did now show a clear skew towards either parent (15%).

Interestingly, allele distributions in progeny originating from the same basidium were often collectively skewed towards the same parent (**Figure 4**). For example, all 13 progeny strains, including strain 98, that originated from basidium 6 contained significantly more serotype D alleles than serotype A alleles, with the number of loci homo-/hemi-zygous for serotype D ranging from 18 (progeny strains 95, 97, 99, 101) to 30 (progeny strain 98). Similarly, in basidium 16, 79% of the 33 genotyped loci across all 14 progeny strains were homo-/hemi-zygous serotype D allele while only 8% contained serotype A alleles: the remaining 13% were heterozygous. In total, 13 out of the 20 investigated basidia were collectively skewed towards *C. deneoformans* alleles. Only three basidia were skewed in favor of the *C. neoformans* parent with their proportions of *C. neoformans* alleles ranging from 50% to 64%. In the remaining four basidia, the majority of the loci were heterozygous (45%, 48%, 49%, 54%).

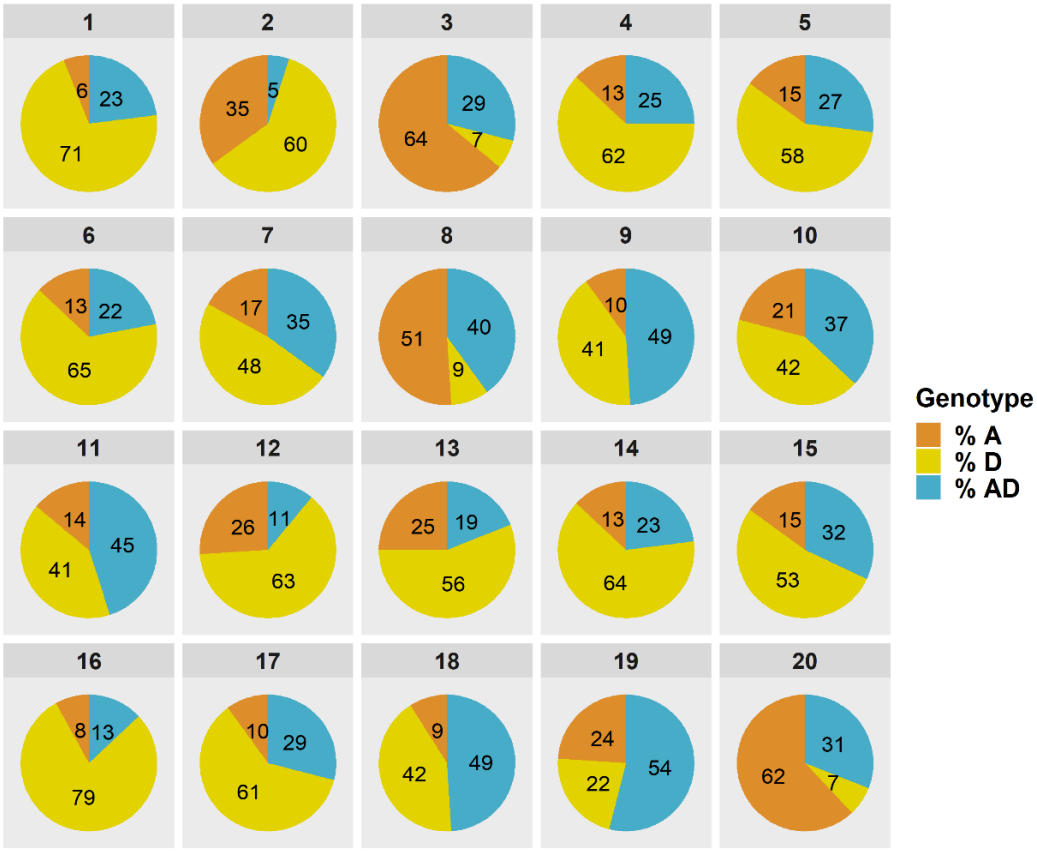


Figure 4: Genotype distribution among hybrid strains originating from different basidia. Genotype data from strains belonging to the same basidia were pooled together and the proportions of loci homo-/hemizygous for serotype A and D alleles, as well as heterozygous loci were calculated. These pie charts display the proportions of the three genotypes found within each basidium.

Preferential retention of parental chromosomes: We conducted Chi-square tests on pooled genotype data for each chromosome to determine if the observed proportions of A, D and AD genotypes deviated significantly from a ratio of 1:1:1. This 1:1:1 ratio is derived based on the assumption that on average, each homologous chromosome pair had one chiasma per chromosome arm per meiotic event and that 50% of the chiasmata would cause chromosomal non-disjunction, resulting in the production of a heterozygote while the two remaining non-sister chromatids segregated properly into the next generation. Overall, our results indicated significant deviations from the 1:1:1 ratio for chromosomes 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, and 13 ($p < 0.05$) (**Figure 5)**. At chromosomes 1, 6, 7, 8 and 13, the progeny population showed a significant preference for the serotype D copy, whereas being heterozygous was preferred at chromosomes 2, 3, 9, and 11. Chromosome 4 was the only one showing statistically significant preference by the progeny from the serotype A parent. The remaining three chromosomes 5, 12 and 14 showed no statistically significant preference towards any of the three categories.

A similar analysis using Chi-square tests was performed to determine if either serotype A or D allele was preferred at each of the 14 chromosomes at the whole progeny population level. In this analysis, all progeny strains were assumed diploid and the null hypothesis was that the A and D alleles were in equal frequency in the population, at 50% each. The results of our analyses separated the 14 chromosomes into three groups. The first group involved chromosomes 1, 2, 6, 7, 8, 10 and 13, with each of the chromosomes showing a preference (i.e. significantly more than 50%) for the D allele in the progeny population ($p<0.05$). The second pattern included three chromosomes 9, 11 and 12 that showed biases in favour of the A allele ($p<0.05$). The remaining four chromosomes (3, 4, 5, and 14) didn't show a significant bias in favour of either the A or the D alleles.

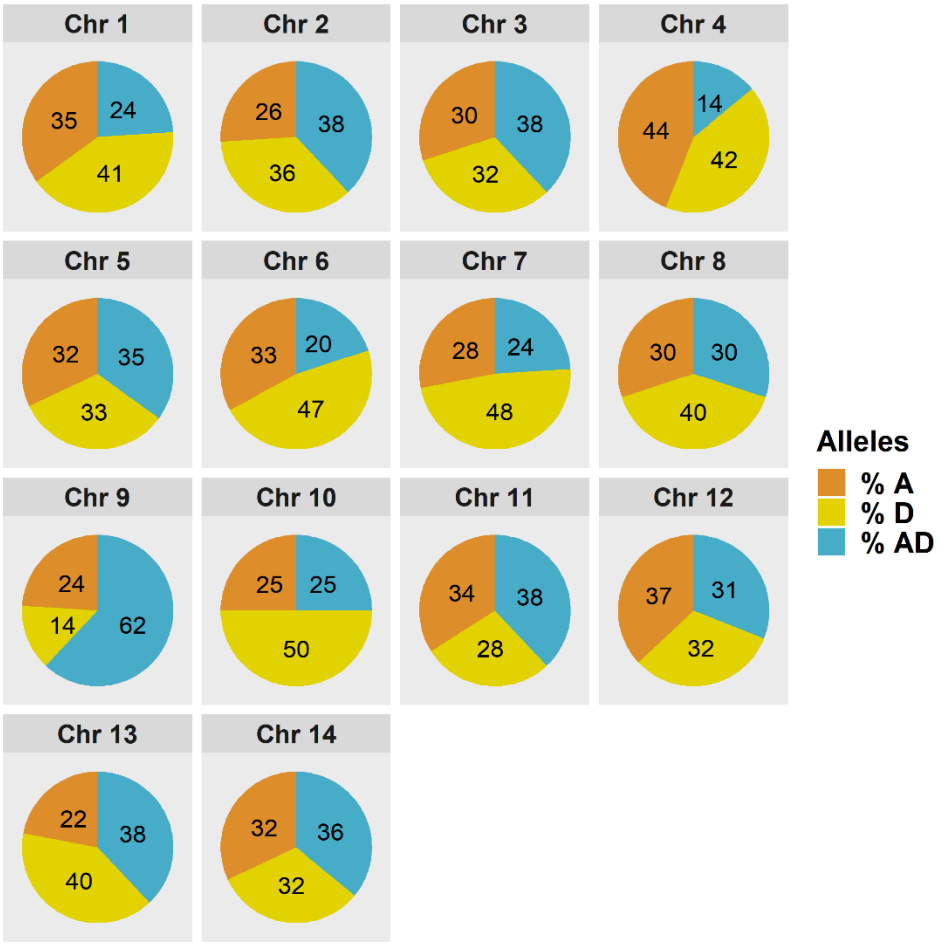


Figure 5: Differential genotype distributions and allele distributions between chromosomes. Genotype data for markers located on the same chromosome were pooled together and the proportions of loci homo-/hemizygous for serotype A and D alleles, as well as heterozygous loci were calculated. These pie charts display the proportions of the three genotypes found at each chromosome.

BDM incompatibilities: A bidirectional BDM incompatibility between markers A6310 and A7310 was detected in our AD hybrid population. None of the 297 hybrid strains were homo-/hemi-zygous for one parental allele at A6310 while being homo/hemizygous for the other parental allele at A7310. All other possible genotype combinations between these two markers were present in the population (A & A, D & D, A & AD, D & AD, AD & AD, AD & A) except for heterozygous at A6310 paired with homo-/hemi-zygous for D allele at A7310 (AD & D).

7.6 Discussion

Hybrids, by definition, are expected to contain a mosaic genome composed of genetic elements from both parents. Indeed, in certain cases, it is possible to use the whole

genome sequence of hybrids to reconstruct the ancestral sub-genomes of the parents[35]. However, AD hybrids surveyed here displayed significantly lower heterozygosity than expected: approximately 67% of their genomes were homo-/hemi-zygous for alleles from a single parent. In contrast, previous studies have found the average heterozygosity of natural AD hybrids to be 60-79%, based on genotypic data from loci located on multiple chromosomes, with AD hybrids containing higher overall heterozygosity showing greater virulence in mouse models[15,17]. It is possible that clinical and environmental settings, as opposed to laboratory conditions, present unique stressors to cryptococcal hybrids where retaining heterozygosity may be advantageous for better survival. Indeed, we note that the average heterozygosity of our study population was even lower than that observed in a previous, lab-derived AD hybrid population with the same parental strains (CDC15 and JEC20). In that study, the hybrid spores were collected by washing and were germinated at 37°C, which, as opposed to the optimum growth temperature of 30°C used in our study, creates thermal stress comparable to a mammalian infection[36]. In addition, the serotype D parent JEC20 is less thermotolerant than the serotype A parent CDC15. Therefore, our results suggest that in relatively stress-free environments such as an optimal temperature and abundant nutrients and moisture, serotype D alleles could be favoured to germinate over serotype A alleles. However, in certain cases such as at chromosome 9, our study population contained a high proportion (62%) of heterozygous loci, suggesting that retaining both parental copies of this chromosome might offer hybrids a fitness advantage.

Consistent with our expectation, AD hybrid progeny from the same basidium showed similar genome-wide distortions in allele distributions. This is also consistent with previous results of mitotic recombination occurring within basidia which creates spores that are collectively biased towards one parental genome[11]. However, in the majority of basidia, a high diversity of nuclear genotypes was found among the spores, often more than expected based on a single round of meiosis (four recombinant haploid nuclei), either with or without random packaging of two nuclei per basidiospore (a maximum of 10 diploid genotypes total per basidium). Given the high degree of genetic diversity, it is likely that hybrids continue to lose

chromosomes, preferentially of the same parent, within the basidium or immediately after germination which could further skews the genome-wide allele distribution in favor of one parent. Incompatibility between the two parental genomes, as evidenced by the presence of a BDM incompatibility detected between two markers on chromosome 1, could also contribute to differential chromosome loss between the two parents. This BDM incompatibility was reported to be unidirectional in a previous study whereas, in our population, both reciprocal genotypes were absent[11]. This inconsistency was likely due to random sampling effects between the two studies.

All progeny from this cross inherited the mitochondrial genome from the serotype D parent JEC20. Thus, it is tempting to speculate that the clear preference of the hybrid progeny for the *C. deneoformans*' nuclear genome could be due to inheriting mitochondria from the same parent. In *Cryptococcus* species, mitochondrial inheritance is uniparental with meiotic progeny inheriting mitochondria from the *MAT***a** parent[37,38]. In our cross, the *C. deneoformans* strain JEC20 was the *MAT***a** parent. While mitochondrial genomes are typically inherited independently from the nuclear genomes, interactions and coordination between the two genomes are essential for proper cellular function. Specifically, key protein complexes in mitochondria contain subunits encoded by both nuclear and mitochondrial genomes[39]. The estimated 10-15% nucleotide sequence divergence translates to the two species being different at ~2.5 million bases, given a total genome length of 20 Mb[40]. Consequently, key nuclear genes required for interactions with the mitochondrial genome are likely polymorphic between *C. neoformans* and *C. deneoformans.* Therefore, AD hybrids that retained components of *MAT***a** parent's nuclear genome could be more likely to germinate due to their increased compatibility with their mitochondrial genomes inherited from the same parent. Similar analysis of progeny derived from a reciprocal hybrid cross between serotype A, *MAT***a** and a serotype D, *MAT*α parents is needed to test this hypothesis.

At the whole progeny population level, seven of the 14 chromosomes showed a biased allele distribution in favour of the serotype D alleles while three showed a skew towards the serotype A alleles, with the remaining four chromosomes showing no significant bias. The overall allelic biases were consistent with the genotype biases

across the genome. However, the chromosome-specific biases suggest differences among chromosomes in spore germination in a stress-free environment between the two parental genomes. While *C. neoformans* alleles are generally associated with higher virulence *in vivo*, their presence might be disadvantageous in the absence of stress, and especially in the light of possible incongruence with the mitochondrial genome.

The germination rate observed in our interspecific cross between *C. neoformans* and *C. deneoformans* is consistent with previous estimates of approximately ~20%, suggesting that the majority of hybrid spores were inviable[5,11,12,33]. However, a significant variation in germination rates was observed among the 31 dissected basidia in our current study. Interestingly, all three basidia whose allele distributions were biased in favor of the *C. neoformans* parent had significantly higher germination rates than other basidia, at 29.3%, 54.3% and 73.5% ($p < 0.05$). This suggests that viability of hybrid progeny is highly sensitive to the complement of parental alleles they receive: if meiosis results in daughter nuclei receiving favorable combinations of parental chromosomes due to limited crossing over and/or co-segregation of chromosomes from the same parent, viability of resulting spores would increase. While our results point to *C. deneoformans* chromosomes as being preferred overall, the high germination rates of basidia skewed in favor of the *C. neoformans* parent suggests that it is possible to generate viable progeny given the right combination of genetic material from either parent.

The increasing prevalence of AD hybrids among clinical samples necessitates a better understanding of the hybridization process between *Cryptococcus* species and the potential for increased virulence in resulting hybrids[7]. Transgressive segregation in virulence and virulence-associated phenotypes has been observed in previous interspecific crosses of *Cryptococcus*[16,20,22]. Our results highlight the significant departure of allele distributions from expected Mendelian ratios in AD hybrids. We also observed high genetic diversity being generated after a single hybridization event, as a result of meiotic recombination and mitotic recombination. The novelty and plasticity of their recombinant genomes could drive the hybrids' evolution in successfully adapting to novel environments and competing with the parental lineages.

7.7 Supplementary Data

Table S1: PCR-RFLP markers used to genotype AD hybrids

Marker ID	Primer Sequence	Location A (bp)	Size A (bp)	Location D (bp)	Size D (bp)	Enzyme
CNA00050F	AGCAATTCCAAACCGACCCC	31347	1440	17738	1436	HaeIII
CNA00050R	TTACGCCCACACCAAGGCAT	29908		19173		SacI
CNA06310F	CGTATCCAGCGGCTGACCAC	1696056	1100	1715148	1105	PvuII
CNA06310R	AATGGTCGCTAGACTGGACGGG	1697155		1716252		
CNA07310F	ACACCCCAAATTCCCCAACC	1973202	975	2000368	975	PstI
CNA07310R	ACATCCCAAACGAACCCAC	1974176		2001342		
CNA07990F	TCCAATGGACGAGGACGATG	2177600	1430	2200900	1424	HinfI
CNA07990R	TGACCGGTGTGGGTTGCAAT	2179029		2202323		
CNB00360F	AGTGCTCAGAGTCTGGGGCTGG	101410	1214	102897	1223	HinclI
CNB00360R	GCCATTCGCAGGGGTGGAGG	102623		104119		
CNB05710F	TTGGACAACGCAAGACCCCAG	1598321	1262	1601790	1263	HindIII
CNB05710R	TGTTGCAAGCAACGATGCC	1599582		1603052		
CNC00670F	TGTGCGGCTTTGGGATTGGT	195225	1110	191360	1108	HaeIII
CNC00670R	TTGCAGGTATGGCCGAATGG	196334		192467		
CNC07180F	TGGAGGCGTTGGGCGAAATAGAG	15726	1316	2100876	1309	HaeIII
CNC07180R	TTCAGCCGTCGCCTTTACCACAA	17041		2102184		
CND00510F	CGGTGCCGCTTTATTTGTGGC	1650153	1325	150300	1327	HinfI
CND00510R	TCTAGCGCCAAAGCGTGCAAG	1651477		151626		
RUM1	TGAAGATTTTGGATTCGAAGAAGGTGACG	258645	1345	59083	1340	EcoRV
RUM1	CAAGTGCAGAGCTGATCGGCATGGG	259989		57744		
CNE00250F	TGGCGTCTCTTTGAACGCGATC	44253	1118	55046	1118	HaeIII
CNE00250R	ATGGCGGAATGTCCGGCTTT	45370		56163		
CNE05150F	CCACACTTCCAACCATCCTCG	1417863	1182	1445761	1185	HaeIII
CNE05150R	CATGACCTTCGGCTCCAGCAC	1416682		1444577		
CNF00290F	TCATGCCCCTTCGCCTTCAT	1307533	1457	96744	1455	HaeIII
CNF00290R	TTCTCCTTCTCCCCATCCCA	1308989		98198		
CNF04800F	GGTCTGCTCTATATGAGCTGCG	49162	1044	1397215	1058	HindIII
CNF04800R	GGAAACTGGTCCCCAACAATC	48119		1398272		
CNG00170F	TTTCTTCCGCCGCTTCTCAC	1348500	1245	42764	1232	HaeIII
CNG00170R	ACAGCGCGTTGAGTTTCGGT	1349744		43995		
CNG03250F	TTGCCAATAACGTGGCACGG	444807	1424	916601	1423	HindIII
CNG03250R	AAAGGGAGGCGGCTGATGATA	446230		918023		
CNG04610F	TGTTTCCCACAGGCCAAGGACT	59766	1466	1308769	1514	HinfI
CNG04610R	CGTGCGCGAATGCATCGATAT	61231		1310282		
CNH00030F	TGTCGATGTGCTTCTCGGCA	12634	1272	1186643	1274	PstI
CNH00030R	CTCCCTCCCATCCCAAAACAC	13905		1187916		
CNH02750F	TTGGATCGCTTGCTCGCGAA	818314	1428	349466	1411	EcoRI
CNH02750R	AGGCCCGAGCAAAGGAATGA	819741		350876		
CNH03360F	CGGGGCTATTTGGAGCGAAA	968260	1396	150788	1380	HaeIII
CNH03360R	ATGATGGGGCCTCTGGATTG	969655		152167		
CNI00070F	CCGCCTGCACACCTTTCTT	1142613	1214	20265	1200	XhoI
CNI00070R	TGTCTTCGGTTTGGATGGG	1143826		21464		
CNI04370F	AGCGGCTACAGCAAAAGCGA	30620	1078	1166708	1084	HinfI
CNI04370R	AACATGTCCGCCTCACCCAA	31697		1167791		
CNJ00070F	ATGGCGGAAGAGGCGTATGA	1016363	1153	14214	1135	HinfI
CNJ00070R	CCTGTCCAGTGCGCATTTCG	1017515		15348		
CNJ02920F	TGGGGGAGAAAGGACATTGG	175236	1555	904610	1568	HinfI
CNJ02920R	CAAATGCCGAGCTCCCTTC	176790		906177		
CNK00170F	AACATGGCATCTCCCCCCAA	1500435	1108	54380	1111	HinfI
CNK00170R	TCGTGCTGACCATGCGGTTT	1501542		55490		
CNK03410F	TTTCGCCGCACCCCCTTTTT	1513810	1131	1002531	1125	HinfI
CNK03410R	CCTCGCCGCCCAATAATTCA	1514940		1003655		
CNL03990F	CCCAGGCAGCCGAGGATG	787025	1497	95393	1487	HaeIII
CNL03990R	GCTCGTCGTGACCAGAGGCG	785529		96879		
CNL04620F	TTCGTGGCGACAGGTTTTGGG	595892	1321	289852	1321	HindIII
CNL04620R	TTCAGCGATGGGTTGAGGCA	597212		291172		
CNL06810F	TTAATGGACTGGGCAGATGCTCGTC	18019	868	894746	879	SacI
CNL06810R	ATGTCTTCTCCCGCCCTTTTTGCC	18886		895624		
CNM00180F	GCTCAAGAACCATACCTGCTCAT	65845	1607	39248	1633	HinclI
CNM00180R	GGCGGCAGGTGACTTCAGTG	67451		40880		
CNM02560F	ATGGACGCTCTCACATTACCTTGC	757906	1491	776201	1493	AccI
CNM02560R	ACGCTGCCCTCTCCCACAGTC	759396		777693		
CNN00060F	CCCAACCTCATCCCACCTC	40214	1223	18763	1228	XhoI
CNN00060R	ACAGAACCCATTGAGCCCGA	41436		19990		
CNN02060F	TTGGAACAGGCCACTCGGAA	632500	1432	644695	1433	HaeIII
CNN02060R	ACCGCCAAGGATTCTTGCGA	633931		646127		

Table S2: The position of hybrid spores that germinated on the spore chains, germination rates and the number of unique multi-locus genotypes found within individual basidia are shown here.

Progeny	Basidia	Position	Spores plated	Spores germinated	Germination rate (%)	Number of unique multi-locus genotypes
1	1	2 - 3	39	8	20.5	6
2		3 - 10				
3		3 - 10				
4		3 - 10				
5		3 - 10				
6		3 - 10				
7		3 - 10				
8		3 - 10				
9	2	1 - 3	31	9	29	5
10		1 - 3				
11		1 - 3				
12		3 - 5				
13		5 - 6				
14		5 - 6				
15		7				
16		7				
17		8				
18	3	1 - 3	68	50	73.5	46
19		1 - 3				
20		1 - 3				
21		1 - 3				
296		1 - 3				
22		3 - 6				
23		3 - 6				
24		3 - 6				
25		3 - 6				
26		3 - 6				
27		3 - 6				
28		3 - 6				
29		3 - 6				
30		3 - 6				
31		6 -7				

32		6 -7				
33		7 - 9				
34		7 - 9				
35		7 - 9				
36		7 - 9				
37		7 - 9				
38		7 - 9				
39		7 - 9				
40		9 - 11				
41		9 - 11				
42		9 - 11				
43		9 - 11				
44		9 - 11				
45		9 - 11				
46		9 - 11				
47		9 - 11				
48		9 - 11				
49		9 - 11				
50		12 - 18				
51		12 - 18				
52		12 - 18				
53		12 - 18				
54		12 - 18				
55		12 - 18				
56		12 -15				
57		12 -15				
58		12 -15				
59		12 -15				
60		12 -15				
61		12 -15				
62		12 -15				
63		15 - 17				
64		15 - 17				
65		15 - 17				
66		15 - 17				
67	4	1 - 4	39	9	23.1	9
68		1 - 4				
69		1 - 4				

70 71 72 73 74 75		1 - 4 1 - 4 4 - 7 4 - 7 8 9 - 10				
76 77 78 79 80 81 82 83 84 85 86 87 88	5	1 - 2 1 - 2 2- 7 2- 7 2- 7 2- 7 2- 7 2- 7 2- 7 2- 7 2- 7 2- 7 2- 7	25	13	52	13
89 90 91 92 93 94 95 96 97 98 99 100 101	6	1 - 2 2 3 - 6 3 - 6 3 - 6 9 - 12 6 - 11 6 - 11 6 - 11 6 - 11 6 - 11 6 - 11 6 - 11	43	13	30.2	13
102 103	7	1 4 - 6	28	2	7.1	2
104 105 106 107	8	1 - 11 1 - 11 1 - 11 1 - 11	58	17	29.3	16

108		1				
109		3				
110		4 - 6				
111		4 - 6				
112		4 - 6				
113		4 - 6				
114		6 - 8				
115		6 - 8				
116		8 - 10				
117		8 - 10				
118		8 - 10				
119		10 - 12				
120		10 - 12				
121	9	1 - 7	69	28	40.6	18
122		1 - 7				
123		1 - 7				
124		1 - 7				
125		1 - 7				
126		1 - 7				
127		1 - 7				
128		1 - 7				
129		1 - 7				
130		1 - 7				
131		1 - 7				
132		7 - 9				
133		7 - 9				
134		7 - 9				
135		7 - 9				
136		7 - 9				
137		10 - 17				
138		10 - 17				
139		10 - 17				
140		10 - 17				
141		10 - 17				
142		10 - 17				
143		10 - 17				
144		10 - 17				
145		10 - 17				

146 147 148		10 - 17 10 - 17 17 - 20				
149 150 151 152 153 154	10	1 - 4 8 - 14 8 - 14 8 - 14 8 - 14 14 - 17	62	6	9.7	6
155 156 157 158 159 160 161	11	1 - 3 4 - 6 1 - 20 1 - 20 1 - 20 6 - 7 11 - 12	52	7	13.5	6
162 163 164 165 166 167 168 169 170	12	1 - 5 1 - 5 1 - 5 1 - 5 5 - 7 5 - 7 5 - 6 5 - 8 9 - 10	32	9	28.1	8
171 172 173	13	6 - 8 6 - 8 13	39	3	7.7	3
174 175 176 177 178 179 180 181 182 183	14	1 - 4 1 - 4 1 - 4 1 - 4 1 - 4 4 - 6 6 - 8 8 - 10 8 - 10 10	71	18	25.4	13

184		10				
185		11 - 13				
186		11 - 13				
187		11 - 13				
188		11 - 13				
189		15 - 19				
190		15 - 19				
191		15 - 19				
192	15	1 - 3	66	19	28.8	19
295		3 - 4				
193		4				
194		4				
195		5 - 6				
196		5 - 6				
197		6 - 9				
198		6 - 9				
199		6 - 9				
200		6 - 9				
201		6 - 9				
202		10 - 14				
203		10 - 14				
204		10 - 14				
205		10 - 14				
206		10 - 14				
207		10 - 14				
208		10 - 14				
209		14 - 16				
210	16	1 - 9	69	14	20.3	10
211		1 - 9				
212		1 - 9				
213		1 - 9				
214		1 - 9				
215		1 - 9				
216		1 - 9				
217		1 - 9				
218		1 - 9				
219		9 - 14				
220		15 – 19				

221 222 223		15 – 19 15 – 19 15 – 19				
224 225 226 227 228	17	5 8 5 - 6 9 - 12 9 - 12	59	5	8.5	5
229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244	18	1 - 3 1 - 3 4 - 5 5 - 6 5 - 6 5 - 6 6 - 10 6 - 10 6 - 10 6 - 10 6 - 10 6 - 10 10 - 16 10 - 16 10 - 16 16 - 17	71	16	22.5	16
245 246 247 248 249 250 251 252 253 254 255 256 257	19	3 - 4 5 - 6 6 - 7 7 - 8 9 - 10 7 - 11 7 - 11 7 - 11 11 - 15 11 - 15 11 - 15 11 - 15 11 - 13	69	13	18.8	10
258	20	1 - 4	70	38	54.3	38

259		1 - 4				
260		1 - 4				
261		1 - 3				
262		1 - 3				
263		1 - 3				
264		3 - 7				
265		3 - 7				
266		3 - 7				
267		3 - 7				
268		3 - 7				
269		3 - 7				
270		3 - 7				
271		3 - 7				
272		3 - 6				
273		3 - 6				
274		7 - 10				
275		7 - 10				
276		7 - 10				
277		7 - 10				
278		7 - 10				
279		7 - 10				
280		7 - 10				
281		7 - 10				
282		11 - 12				
283		11 - 12				
284		11 - 12				
285		11 - 12				
286		13 - 16				
287		13 - 16				
288		13 - 16				
289		13 - 16				
290		13 - 16				
291		13 - 16				
292		13 - 16				
293		13 - 16				
294		13 - 16				
297		13 - 16				

7.8 References

1. Rajasingham R, Smith RM, Park BJ, et al. Global burden of disease of HIV-associated cryptococcal meningitis: an updated analysis. *Lancet Infect Dis*. 2017;17(8):873-881. doi:10.1016/S1473-3099(17)30243-8

2. Litvintseva AP, Xu J, Mitchell TG. *Cryptococcus*: from human pathogen to model yeast. In: Heitman J, Kozel TR, Kwon-Chung KJ, Perfect JR, Casadevall A, eds. *Cryptococcus: From Human Pathogen to Model Yeast.* ASM Press; 2010:97-112.

3. Cogliati M. Global Molecular Epidemiology of *Cryptococcus neoformans* and *Cryptococcus gattii*: An Atlas of the Molecular Types. *Scientifica (Cairo)*. 2013;2013. doi:10.1155/2013/675213

4. Tanaka R, Nishimura K, Miyaji M. Ploidy of serotype AD strains of *Cryptococcus neoformans*. *Japanese J Med Mycol*. 1999;40(1):31-34. http://www.ncbi.nlm.nih.gov/pubmed/9929580. Accessed February 28, 2018.

5. Lengeler KB, Cox GM, Heitman J. Serotype AD strains of *Cryptococcus neoformans* are diploid or aneuploid and are heterozygous at the mating-type locus. *Infect Immun*. 2001;69(1):115-122. doi:10.1128/IAI.69.1.115-122.2001

6. Bennett JE, Kwon-Chung KJ, Howard DH. Epidemiologic differences among serotypes of *Cryptococcus neoformans*. *Am J Epidemiol*. 1977;105(6):582-586. http://www.ncbi.nlm.nih.gov/pubmed/326036.

7. Samarasinghe H, Xu J. Hybrids and hybridization in the *Cryptococcus neoformans* and *Cryptococcus gattii* species complexes. *Infect Genet Evol*. 2018;66:245-255. doi:10.1016/J.MEEGID.2018.10.011

8. Maduro AP, Mansinho K, Teles F, et al. Insights on the Genotype Distribution Among *Cryptococcus neoformans* and C. *gattii* Portuguese Clinical Isolates. *Curr Microbiol*. 2014;68(2):199-203. doi:10.1007/s00284-013-0452-0

9. Cogliati M, Esposto MC, Clarke DL, Wickes BL, Viviani MA. Origin of *Cryptococcus neoformans* var. *neoformans* diploid strains. *J Clin Microbiol*. 2001;39(11):3889-3894. doi:10.1128/JCM.39.11.3889-3894.2001

10. Desnos-Ollivier M, Patel S, Raoux-Barbot D, Heitman J, Dromer F. Cryptococcosis Serotypes Impact Outcome and Provide Evidence of *Cryptococcus neoformans* Speciation. *MBio*. 2015;6(3):e00311-15. doi:10.1128/mBio.00311-15

11. Vogan AA, Khankhet J, Xu J. Evidence for Mitotic Recombination within the Basidia of a Hybrid Cross of *Cryptococcus neoformans*. *PLoS One*. 2013;8(5). doi:10.1371/journal.pone.0062790

12. Forsythe A, Vogan A, Xu J. Genetic and environmental influences on the germination of basidiospores in the *Cryptococcus neoformans* species complex. *Sci Rep*. 2016;6:33828. doi:10.1038/srep33828

13. Sun S, Xu J. Genetic analyses of a hybrid cross between serotypes A and D strains of the human pathogenic fungus *Cryptococcus neoformans*. *Genetics*. 2007;177(3):1475-1486. doi:10.1534/genetics.107.078923

14. Tanaka R, Nishimura K, Imanishi Y, Takahashi I, Hata Y, Miyaji M. Analysis of Serotype AD Strains from F1 Progenies between Urease-positive- and Negative-strains of *Cryptococcus neoformans*. *Japanese J Med Mycol*. 2003. doi:10.3314/jjmm.44.293

15. Cogliati M, Barchiesi F, Spreghini E, Tortorano AM. Heterozygosis and Pathogenicity of *Cryptococcus neoformans* AD-Hybrid Isolates. *Mycopathologia*. 2011;173(5-6):347-357. doi:10.1007/s11046-011-9467-x

16. Vogan AA, Khankhet J, Samarasinghe H, Xu J. Identification of QTLs Associated with Virulence Related Traits and Drug Resistance in *Cryptococcus neoformans*. *G3 (Bethesda)*. 2016;6(9):2745–2759. doi:10.1534/G3.116.029595

17. Aminnejad M, Cogliati M, Duan S, et al. Identification and Characterization of VNI/VNII and Novel VNII/VNIV Hybrids and Impact of Hybridization on Virulence and Antifungal Susceptibility Within the C. *neoformans*/C. *gattii* Species Complex. *PLoS One*. 2016;11(10):e0163955. doi:10.1371/journal.pone.0163955

18. Chaturvedi V, Fan J, Stein B, et al. Molecular genetic analyses of mating pheromones reveal intervariety mating or hybridization in *Cryptococcus neoformans*. *Infect Immun*. 2002;70(9):5225-5235. doi:10.1128/IAI.70.9.5225-5235.2002

19. Lin X, Litvintseva AP, Nielsen K, et al. αADα Hybrids of *Cryptococcus neoformans*: Evidence of Same-Sex Mating in Nature and Hybrid Fitness. *PLoS Genet*. 2007;3(10):e186. doi:10.1371/journal.pgen.0030186

20. Litvintseva AP, Lin X, Templeton I, Heitman J, Mitchell TG. Many globally isolated AD hybrid strains of *Cryptococcus neoformans* originated in Africa. *PLoS Pathog*. 2007;3(8):1109-1117. doi:10.1371/journal.ppat.0030114

21. Li M, Liao Y, Chen M, Pan W, Weng L. Antifungal susceptibilities of *Cryptococcus* species complex isolates from AIDS and non-AIDS patients in Southeast China. *Brazilian J Infect Dis*. 2012. doi:10.1016/S1413-8670(12)70301-X

22. Li W, Averette AF, Desnos-Ollivier M, Ni M, Dromer F, Heitman J. Genetic Diversity and Genomic Plasticity of *Cryptococcus neoformans* AD Hybrid Strains. *G3 (Bethesda)*. 2012;2(1):83-97. doi:10.1534/g3.111.001255

23. Hu G, Liu I, Sham A, Stajich JE, Dietrich FS, Kronstad JW. Comparative hybridization reveals extensive genome variation in the AIDS-associated pathogen *Cryptococcus neoformans*. *Genome Biol*. 2008;9(2):R41. doi:10.1186/gb-2008-9-2-r41

24. Cuomo CA, Rhodes J, Desjardins CA. Advances in *Cryptococcus* genomics: insights into the evolution of pathogenesis. *Mem Inst Oswaldo Cruz*. 2018;113(7):e170473. doi:10.1590/0074-02760170473

25. Rhodes J, Desjardins CA, Sykes SM, et al. Population Genomics Of *Cryptococcus neoformans* var. *grubii* Reveals New Biogeographic Relationships And Finely Maps Hybridization. *bioRxiv*. May 2017:132894. doi:10.1101/132894

26. Xu J, Vilgalys R, Mitchell TG. Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*. *Mol Ecol*. 2000;9(10):1471-1481. doi:10.1046/j.1365-294x.2000.01021.x

27. D'Souza CA, Kronstad JW, Taylor G, et al. Genome variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts. *MBio*. 2011;2(1):e00342-10. doi:10.1128/mBio.00342-10

28. Hagen F, Khayhan K, Theelen B, et al. Recognition of seven species in the *Cryptococcus gattii*/*Cryptococcus neoformans* species complex. *Fungal Genet Biol*. 2015;78:16-48. doi:10.1016/j.fgb.2015.02.009

29. Kwon-Chung KJ, Bennett JE, Wickes BL, et al. The Case for Adopting the "Species Complex" Nomenclature for the Etiologic Agents of Cryptococcosis. *mSphere*. 2017;2(1):e00357-16. http://msphere.asm.org/content/2/1/e00357-16#sec-11. Accessed July 22, 2017.

30. Hagen F, Lumbsch HT, Arsic Arsenijevic V, et al. Importance of Resolving Fungal Nomenclature: the Case of Multiple Pathogenic Species in the *Cryptococcus* Genus. *mSphere*. 2017;2(4):e00238-17. doi:10.1128/mSphere.00238-17

31. Dobzhansky T. Studies on Hybrid Sterility. II. Localization of Sterility Factors in Drosophila Pseudoobscura Hybrids. *Genetics*. 1936;21(2):113-135. http://www.ncbi.nlm.nih.gov/pubmed/17246786. Accessed July 3, 2019.

32. Orr HA. Dobzhansky, Bateson, and the Genetics of Speciation. *Genetics*. 1996;144(4):1331-1335.

33. Vogan AA, Xu J. Evidence for genetic incompatibilities associated with post-zygotic reproductive isolation in the human fungal pathogen *Cryptococcus neoformans*. *Genome*. 2014;57(6):335-344. doi:10.1139/gen-2014-0077

34. Ellis D, Pfeiffer T. Ecology, life cycle, and infectious propagule of *Cryptococcus neoformans*. *Lancet*. 1990;336(8720):923-925. doi:10.1016/0140-6736(90)92283-N

35. Louis VL, Despons L, Friedrich A, et al. Pichia sorbitophila, an Interspecies Yeast Hybrid, Reveals Early Steps of Genome Resolution After Polyploidization. *G3 (Bethesda)*. 2012;2(2):299-311. doi:10.1534/g3.111.000745

36. Sun S, Xu J. Genetic analyses of a hybrid cross between serotypes A and D strains of the human pathogenic fungus *Cryptococcus neoformans*. *Genetics*. 2007;177(3):1475-1486. doi:10.1534/genetics.107.078923

37. Xu J, Ali RY, Gregory DA, et al. Uniparental Mitochondrial Transmission in Sexual Crosses in *Cryptococcus neoformans*. *Curr Microbiol*. 2000;40(4):269-273. doi:10.1007/s002849910053

38. Yan Z, Xu J. Mitochondria Are Inherited From the MATa Parent in Crosses of the Basidiomycete Fungus *Cryptococcus neoformans*. *Genetics*. 2003;163(4):1315-1325.

39. Xu J, He L. Current perspectives on mitochondrial inheritance in fungi. *Cell Health Cytoskelet*. 2015;2015(7):143. doi:10.2147/CHC.S59508

40. Kavanaugh LA, Fraser JA, Dietrich FS. Recent Evolution of the Human Pathogen *Cryptococcus neoformans* by Intervarietal Transfer of a 14-Gene Fragment. *Mol Biol Evol*. 2006;23(10):1879-1890. doi:10.1093/molbev/msl070

41. Payseur BA, Rieseberg LH. A genomic perspective on hybridization and speciation. *Mol Ecol*. 2016;25(11):2337-2360. doi:10.1111/mec.13557

42. Coyne JA, Allen Orr H. The evolutionary genetics of speciation. Magurran AE, May RM, eds. *Philos Trans R Soc London Ser B Biol Sci*. 1998;353(1366):287-305. doi:10.1098/rstb.1998.0210

43. Rieseberg LH. Crossing relationships among ancient and experimental sunflower hybrid lineages. *Evolution (N Y)*. 2000;54(3):859-865. doi:10.1111/j.0014-3820.2000.tb00086.x

44. Buerkle CA, Rieseberg LH. The rate of genome stabilization in homoploid hybrid species. *Evolution*. 2008;62(2):266-275. doi:10.1111/j.1558-5646.2007.00267.x

Chapter 8

Conclusions

8.1 Conclusions and Perspectives

This body of work that constitutes my PhD thesis makes several contributions to expanding our knowledge on yeast biology, including their global distribution in soil and the genetics of the infectious yeasts, *C. neoformans* and *C. deneoformans.*

The association between rainfall and biodiversity is well documented in soil microbial communities which is thought to be mediated via the soil organic carbon content (Maestre et al. 2015). Previous metagenomics studies have identified precipitation and soil moisture as determinants of global soil fungal diversity (Egidi et al. 2019; Tedersoo et al. 2014). Due to the lack of a yeast-specific signature in the fungal barcoding gene, high throughput sequencing approaches failed to separate yeasts from other fungi, making it unclear the extent to which overall conclusions apply to this small group of organisms that make up less than 1% of the fungal kingdom. In chapter 2, I identified a significant positive correlation between mean annual precipitation and culturable soil yeast diversity worldwide. A key question is how richness, structure and composition of soil microbial communities would respond to altered precipitation patterns brought on by climate change. Researchers have already linked changes in biomass, stability, biotic interactions, and carbon utilization of soil microbial communities to alterations in precipitation (Hu et al. 2020; Maestre et al. 2015; De Vries and Shade 2013; Zhou, Wang, and Luo 2018). Microbial communities are incredibly complex and any responses to climate change and associated drivers are likely to be multifaceted. Temporal investigations in different biomes that experience perturbed rainfall patterns to different extents are needed to identify how soil yeasts worldwide will be affected by altered precipitation patterns and extreme rainfall or drought events.

Another anthropogenic activity with high impact on ecosystems, specifically on microbes and infectious diseases, is international travel. In chapter 3, I investigated the genetic structure of a largely clonal, novel *C. deneoformans* population discovered in Saudi Arabian desert soil. Its clonal nature points to it having been a recent introduction to the region, rather than being an ancestral population of *C. deneoformans.* In the literature reviews presented in chapters 5 and 6, I summarized the evidence that points to AD hybrids (*C. neoformans* x *C. deneoformans*) arising from a few, independent hybridization events followed by global dispersal, likely facilitated by intercontinental human movement. Following the breakup of the supercontinent Pangea, *C. deneoformans* is believed to have evolved in the Mediterranean Europe while its sister species *C. neoformans* continued to diverge and evolve in Africa (Casadevall et al. 2017). The existence and increasing prevalence of AD hybrids suggest that the two previously isolated sister species are being brought back into contact. In chapter 2, anthropogenic travel was directly implicated when a significant correlation was found between volume of air travel and the number of shared yeast species between countries.

International travel has grown at an unprecedented rate over the last century with the current COVID-19 pandemic being a testament to the crucial role played by human mobility in the global spread of infectious diseases. In recent history, several fungal pathogens that have dealt devastating blows to host populations have been aided by human movement in their global spread. *Batrachochytrium dendrobatidis*, the causative fungus of chytridiomycosis, the worst panzoonotic to date, is believed to have originated in the Korean peninsula, with its spread coinciding with the global expansion of commercial trades in amphibians (O'Hanlon et al. 2018). Human migration has been posited as a possible explanation for the rapid, global spread of the multidrug-resistant, superbug yeast *Candida auris* (Chow et al. 2020). Findings of chapter 2 reaffirmed soil as a primary reservoir of pathogenic yeasts such as *Cryptococcus* and other *Candida* species, which are also capable of benefiting from anthropogenic influences on their habitats. Alternations in their natural habitats could lead to them outgrowing their competition to become threats to global public health. Phylogeographic analyses using whole genome sequences could provide more

conclusive answers on the origin and spread of these yeasts. There is a distinct lack of whole genome sequences available for most of these yeast pathogens: currently, only three, full *C. deneoformans* genomes are publicly available on NCBI. Future research would significantly benefit from investing more on whole genome sequences of yeast pathogens which could provide decision makers with compelling evidence to act before they grow into global threats.

Hybridization is a widespread phenomenon in all domains of life and a driving force in the evolution of species. With anthropogenic activity playing a major role in reacquainting previously separated species, hybridization and resulting hybrid progeny are becoming more apparent in the environment. Outcomes of hybridization primarily depend on the genetic differentiation of parental genomes and resulting genomic, cellular and/or physiological incompatibilities. In chapter 7, I provided evidence for biased, genome-wide allele distributions in AD hybrids, which point to each parent having co-adapted alleles that are no longer compatible with those from the other parent. This supports the classification of the parents into two separate species, rather than varieties of one species complex (Hagen et al. 2015; Kwon-Chung et al. 2017). Previous studies on virulence of AD hybrids have reported conflicting results with some reporting hybrid vigour while others find hybrids to underperform compared to parents (reviewed in chapters 5 and 6). As shown in chapter 7, a single *in vitro* mating event was capable of generating an incredible amount of genetic diversity among progeny strains. A previous study conducted in our lab found a wide, quantitative range in melanin production, capsule production and thermotolerance in a lab-derived AD hybrid population with some displaying transgressive phenotypes (Vogan et al. 2016). Quantifying the expression of virulence factors, as well as measuring virulence using mouse models or waxworm models, of our AD hybrid population would provide a more complete picture of the mechanisms and outcomes of hybridization between *C. neoformans* and *C. deneoformans*.

The intricate genetic networks underlying the virulence phenotypes in *C. neoformans* and *C. deneoformans* are key contributors to the wide quantitative range of phenotypes observed among their hybrid progeny. In chapter 4, I identified six quantitative trait loci (QTL) contributing to variation in melanin production among a

lab-derived F1 population of *C. deneoformans*. While previous studies have identified genes involved in melanin synthesis and its regulation, only two have attempted to identify genes contributing to quantitative melanin variance observed among strains (Roth et al. 2021; Vogan et al. 2016). The gene encoding *LAC1*, the one enzyme that is essential for melanin biosynthesis, was not identified as a melanin QTL in our study. A *LAC1* allele swap incorporating three amino acid changes between phenotypically dissimilar strains also failed to alter melanin production in the transformed strain. However, each of the six melanin QTLs we identified showed significant differences in expression between the two parental strains used to generate the F1 population. Taken together, our findings suggest that different genes play key roles in determining whether melanin is produced, and how much melanin is produced in *C. deneoformans.* Gene knockout experiments of the six QTLs should be conducted to determine their role in melanin synthesis and any other phenotypes. The wide range of melanin phenotypes observed among the F1 population suggests that many more QTLs remain to be identified. Backcrossing select progeny to parents for several generations prior to whole genome sequencing can eliminate some of the background genetic diversity and help to narrow down the number of potential loci.

As discussed in chapter 4, we were the first group to apply the bulk segregant analysis (BSA) approach in a *Cryptococcus* species. BSA is commonly used in plant and *Saccharomyces cerevisiae* research in identifying QTLs contributing to phenotypic variance of industrial and pathogenic significance (Phadke et al. 2018; Wang et al. 2019; Wilkening et al. 2014; Yang et al. 2021). While whole genome sequencing of individual strains can generate a large amount of useful data, it is often more costly and requires more computational power and bioinformatics skills for effective execution. Pooling DNA of strains with similar, extreme phenotypes, as done in BSA, can in fact aid in highlighting contributory genes while masking background genetic noise. The protocol I developed to perform BSA and my findings from chapter 4 can guide and inform future investigations using BSA in *C. deneoformans.*

Overall, my PhD thesis provided a broad overview of global soil yeast epidemiology before delving deeper into the genetics and virulence phenotypes of the infectious yeast, *C. deneoformans.* My work adds to our existing knowledge of yeast biology

while providing novel insights into how yeast communities and infectious yeasts might respond to ongoing changes in their environments that are largely driven by anthropogenic activities. My work has identified evidence gaps and laid the foundation for future investigations to advance our understanding of global yeast diversity, hybridization, molecular genetics, and virulence.

8.2 References

Casadevall, Arturo, Joudeh B Freij, Christopher Hann-Soden, and John Taylor. 2017. "Continental Drift and Speciation of the Cryptococcus Neoformans and Cryptococcus Gattii Species Complexes." *mSphere* 2(2): e00103-17. http://www.ncbi.nlm.nih.gov/pubmed/28435888 (April 14, 2018).

Chow, Nancy A. et al. 2020. "Tracing the Evolutionary History and Global Expansion of Candida Auris Using Population Genomic Analyses." *mBio* 11(2).

Egidi, Eleonora et al. 2019. "A Few Ascomycota Taxa Dominate Soil Fungal Communities Worldwide." *Nature Communications* 10(1): 1–9. https://doi.org/10.1038/s41467-019-10373-z (March 22, 2021).

Hagen, Ferry et al. 2015. "Recognition of Seven Species in the Cryptococcus Gattii/Cryptococcus Neoformans Species Complex." *Fungal Genetics and Biology* 78: 16–48.

Hu, Yilun et al. 2020. "Effect of Increasing Precipitation and Warming on Microbial Community in Tibetan Alpine Steppe." *Environmental Research* 189: 109917.

Kwon-Chung, Kyung J. et al. 2017. "The Case for Adopting the 'Species Complex' Nomenclature for the Etiologic Agents of Cryptococcosis." *mSphere* 2(1): e00357-16. http://msphere.asm.org/content/2/1/e00357-16#sec-11 (July 22, 2017).

Maestre, Fernando T. et al. 2015. "Increasing Aridity Reduces Soil Microbial Diversity and Abundance in Global Drylands." *Proceedings of the National Academy of Sciences* 112(51): 15684–89. https://www.pnas.org/content/112/51/15684 (August 1, 2021).

O'Hanlon, Simon J. et al. 2018. "Recent Asian Origin of Chytrid Fungi Causing

Global Amphibian Declines." *Science*.

Phadke, Sujal S. et al. 2018. "Genome-Wide Screen for Saccharomyces Cerevisiae Genes Contributing to Opportunistic Pathogenicity in an Invertebrate Model Host." *G3: Genes|Genomes|Genetics* 8(1): 63. /pmc/articles/PMC5765367/ (August 1, 2021).

Roth, Cullen et al. 2021. "Pleiotropy and Epistasis within and between Signaling Pathways Defines the Genetic Architecture of Fungal Virulence." *PLOS Genetics* 17(1): e1009313. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1009313 (July 6, 2021).

Tedersoo, Leho et al. 2014. "Global Diversity and Geography of Soil Fungi." *Science* 346(6213). http://science.sciencemag.org/content/346/6213/1256688 (July 2, 2017).

Vogan, Aaron A., Jordan Khankhet, Himeshi Samarasinghe, and Jianping Xu. 2016. "Identification of QTLs Associated with Virulence Related Traits and Drug Resistance in Cryptococcus Neoformans." *G3 (Bethesda, Md.)* 6(9): 2745–2759.

De Vries, Franciska T, and Ashley Shade. 2013. "Controls on Soil Microbial Community Stability under Climate Change." *Frontiers in Microbiology* 0(SEP): 265.

Wang, Zhen et al. 2019. "QTL Analysis Reveals Genomic Variants Linked to High-Temperature Fermentation Performance in the Industrial Yeast." *Biotechnology for Biofuels* 12(1). /pmc/articles/PMC6423876/ (August 1, 2021).

Wilkening, Stefan et al. 2014. "An Evaluation of High-Throughput Approaches to QTL Mapping in Saccharomyces Cerevisiae." *Genetics* 196(3): 853. /pmc/articles/PMC3948811/ (August 1, 2021).

Yang, Luomiao et al. 2021. "Identification of Candidate Genes Conferring Cold Tolerance to Rice (Oryza Sativa L.) at the Bud-Bursting Stage Using Bulk Segregant Analysis Sequencing and Linkage Mapping." *Frontiers in Plant Science* 0: 402.

Zhou, Zhenghu, Chuankuan Wang, and Yiqi Luo. 2018. "Response of Soil Microbial Communities to Altered Precipitation: A Global Synthesis." *Global Ecology and Biogeography* 27(9): 1121–36. https://onlinelibrary.wiley.com/doi/full/10.1111/geb.12761 (August 1, 2021).